DOCUMENT RESUME

ED 299 143 SE 049 668

AUTHOR	Knapp, Michael S.; And Others
TITLE	Designing and Organizing Assessment in the National Science Foundation. An Approach to Assessing Initiatives in Science Education: Volume 1.
INSTITUTION	SRI International, Menlo Park, Calif.
SPONS AGENCY	National Science Foundation, Washington, D.C. Directorate for Science and Engineering Education.
PUB DATE	Apr 88
CONTRACT	NSF-SPA-8651540
NOTE	111p.; For volume 2 see SE 049 669, summary report see SE 049 670.
AVAILABLE FROM	SRI International, 333 Ravenswood Ave., Room B-S142, Menlo Park, CA 94025 ($10.00).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Case Studies; Cost Effectiveness; Decision Making; Educational Assessment; Educational Finance; *Elementary School Science; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Evaluation Utilization; *Financial Policy; Financial Support; *Foundation Programs; Grants; Needs Assessment; School Support; Science Education; *Secondary School Science
IDENTIFIERS	*National Science Foundation

ABSTRACT

This report presents recommendations to the National Science Foundation (NSF) to guide it in assessing its initiatives in science education. The report outlines appropriate goals, procedures, arrangements, and resources necessary to establish an effective set of assessment practices that build on existing assessment activities in the Foundation, fit with agency culture and constraints, and are both comprehensive and practical. Part 1 of this report presents arguments for improving the Foundation's approach to the assessment of science education initiatives. All of the recommendations to the Foundation appear in this section. Part 2 details two sets of design considerations. The first discusses the framing of assessment questions from different perspectives and the second describes design options and elaborates the strengths and weaknesses of different procedures and mechanisms for carrying out assessments. Part 3 reviews the results of SRI's pilot test of short-term, focused assessment procedures in informal science education. Described are the procedures used, the findings, and methodological lessons for further application of these procedures. (CW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299143

AN APPROACH TO ASSESSING INITIATIVES IN SCIENCE EDUCATION

Volume 1: Designing and Organizing Assessment in the National Science Foundation

April 1988

Prepared for:

THE NATIONAL SCIENCE FOUNDATION

NSF Contract No. SPA-8651540
SRI Project No. 1809

Prepared by:

Michael S. Knapp
Andrew A. Zucker
Mark St. John
Patrick M. Shields
Marian S. Stearns

With the assistance of:

Teresa Middleton
Debra M. Shaver
Dorothy Stewart

SRI International

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael S Knapp
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE 049 628

This report presents the conclusions from the second phase of SRI's "Assessment of Initiatives Available to the National Science Foundation (NSF) in Science Education." Complementing an earlier phase of work, in which SRI discussed opportunities for the Foundation to invest strategically in K-12 science education, the second phase concentrated on ways for NSF to assess its support for science education on an ongoing basis. Both phases are part of the Foundation's response to a congressional mandate that it seek outside assistance in developing its plans and approach to managing its investments in science education.

This volume includes three parts that discuss (1) the approach to assessment, (2) detailed design considerations, and (3) the methodological lessons learned from a pilot test of short-term focused assessments in one area of investment (informal science education). The first of these three parts also exists as a separately bound *Summary Report*. Readers wishing more detail on the results of the Phase II pilot test are referred to *Volume 2: Pilot Assessments of the National Science Foundation's Investments in Informal Science Education*, which includes complete write-ups of the findings from six pilot assessments of NSF's investments in informal science education.

The results of Phase I are reported in the following three volumes:

- The *Summary Report* reviews all findings and conclusions regarding NSF's mission in K-12 science education, the opportunities for the Foundation to make a significant contribution to this level of education, and how NSF can approach these opportunities more strategically.
- *Volume 1: Problems and Opportunities* presents full discussions of NSF's mission, the problems in K-12 science education that are susceptible to NSF's influence, and the opportunities to address these problems.
- *Volume 2: Groundwork for Strategic Investment* contains extended discussions of (1) NSF's "core" functions in science education (promoting professional interchange, generating information and knowledge about science education, and supporting innovation), and (2) the basis for strategic investment. This volume also includes a discussion of study methods, a summary of NSF's 30-year history of funding in K-12 science education, and three commissioned papers (regarding NSF's role in mathematics education, computer science education, and efforts to serve minority students in science).

Any of the above volumes may be requested (at the cost of printing) from SRI International, Room B-S142, 333 Ravenswood Avenue, Menlo Park, CA 94025. ATT: Carolyn Estey. Telephone (415) 859-5109.

The conclusions of this report are those of the authors and contractors and do not necessarily reflect the views of the National Science Foundation or any other agency of government.

HIGHLIGHTS OF THE REPORT

Assessment in Relation to NSF's Science Education Initiatives

In supporting science education, the National Science Foundation (NSF) is, in some instances, funding the enrichment experiences of individual students, but more often it is supporting efforts to improve complex, decentralized education systems. The Foundation's best chance for success lies in a grant support strategy that targets NSF's resources on aspects of these systems that are most susceptible to change and appropriately addressed by federal agencies.

Assessment is a critical part of a proactive funding strategy. The Foundation needs to know what it is supporting and accomplishing--or likely to accomplish--and why, when it invests funds in science education. This information contributes to the Foundation's own planning and good management, and also helps demonstrate to external audiences what NSF is doing for science education. **To serve these needs, we define "assessment" more broadly than conventional forms of program evaluation to include any systematic efforts to inform decisionmaking in NSF by gathering, interpreting, and reporting evidence of various kinds.**

Improving Assessment Practices Within the Foundation

This conception of assessment implies the following focus, procedures, and mechanisms for assessment of science education initiatives in NSF. Building on the steps it has already taken to assess its support for science education, the Foundation should:

- *Refocus assessment activities.* Assessment at all levels should focus on (1) what actually happens as a result of NSF's investments, and (2) the logic, assumptions, and rationale underlying these investments. The Foundation should increase the emphasis on assessing *initiatives* within and across programs, rather than on assessing each grantee's project separately or assessing each grant program taken as a whole (except where the "program" is, in effect, a single initiative).

- *Use procedures and mechanisms that yield a "mosaic" of evidence about initiatives.* Because the Foundation's initiatives in science education are complex, assessment of them should develop evidence from three kinds of sources:

    (1) *Comprehensive assessment studies,* such as several contracted studies now under way within the Directorate for Science and Engineering Education (SEE).

(2) *Documentation activities,* such as grants to document particular projects or initiatives, and data collection systems that assemble descriptive information from grantees on an ongoing basis.

(3) *Short-term special-focus assessment activities,* such as quick case studies, analyses of existing data, and working seminars of experts with particular expertise in the assessment of science education.

NSF has some limited experience with the latter two types of activities, but needs to put an array of mechanisms in place to support these activities on a routine basis (e.g., adjunct staff, task ordering arrangements focused on assessment in science education).

- *Change the approach to project-level assessment.* The current requirement that most grantees deliver to the Foundation a self-assessment of their own projects should be dropped, because it does not produce what NSF needs to answer its own assessment questions, nor does it serve the needs of these projects. Grantees should be encouraged and helped, however, to assess their own projects with "formative" purposes in mind--that is, to gather data that helps them reflect on what they are doing and make mid-course corrections. In addition, grantees should be helped to furnish the Foundation with basic descriptive information about their projects (e.g., as part of the data collection systems referred to above).

Making Assessment Part of Foundation Routine

To make this kind of assessment a part of Foundation routine requires the right roles and locus of control, appropriate incentives and rewards, and sufficient resources.

- *Roles and locus of control.* Managers and staff at each organizational level (e.g., program officer, division director, assistant director) should help set assessment agendas and interpret results; they should also sponsor assessment activities (e.g., through program grant funds) or otherwise arrange for these activities to be undertaken. Centralized assessment units like SEE's Office of Studies and Program Assessment (OSPA) should provide technical support (as OSPA now does), as well as carry out some assessment studies.

- *Incentives and rewards.* Incentives at all levels should be strengthened by restructuring assignments of managers and staff to permit more time for assessment and by rewarding individuals and organizational units for conducting and using assessments. A "climate of support" for assessment should be built within the Foundation as a whole and within directorates.

- *Resources.* Sufficient funds should be allocated to the assessment function--in the range of 2% to 5% of total expenditures for science education. These funds should be dispersed among the budgets of programs, divisions, and specialized units responsible for assessment. These amounts do not necessarily imply an increase in funding for science education; these resources should instead be viewed as an integral part of programmatic support for science education, no matter what the level of funding.

Reasons and Prospects for Improvement

There are compelling reasons for NSF to improve its assessment of initiatives in science education. The Foundation has much to gain by making these improvements, and much to lose through inaction.

- *Internal and external pressures for improvement of assessment are strong.* In addition to its own need for better data and analysis to inform strategic grantmaking in science education, important external bodies--e.g., Congress, the Office of Management and Budget--have called for better assessments of science education funding. The Foundation has yet to develop practices that adequately answer its own or others' questions.

- *There are important consequences of inaction.* The neglect of assessment may lead to unfortunate consequences other than less effective operation. NSF will be open to criticism that it is not managing its funds responsibly, it may have greater difficulty justifying its funding for science education, and it may have unwanted assessments imposed on it.

The groundwork for improving assessment has been laid. For example, SEE's Office of Studies and Program Assessment has been established with a significant budget. SEE has initiated several contracted evaluations of particular science education programs and initiatives (e.g., the College Science Instrumentation Program). NSF has begun to overhaul its Management Information System (MIS), which can help to develop better descriptive documentation on projects supported. In addition, NSF has recently begun new assessment activities outside of education, for example, by establishing an evaluation component in such complex initiatives as the Industry-University Collaborative Research Centers, which provide models that may be used to examine science education investments.

By building on these beginnings, the Foundation has the opportunity to put in place a sophisticated approach to assessing its initiatives that will help to focus and sustain its strategies for improving science education over the long term.

CONTENTS

Highlights of the Report i

PART ONE: AN APPROACH TO ASSESSING SCIENCE EDUCATION INITIATIVES IN THE NATIONAL SCIENCE FOUNDATION 1

Scope of the Report 1

What We Mean by "Assessment" 2

Organization of the Report 2

I THE SPECIAL CHALLENGE OF ASSESSING INITIATIVES IN SCIENCE EDUCATION 3

Forces for Improvement in the Foundation's Assessment Practices 3

Groundwork for Improving Assessment of Science Education Initiatives 5

II ASSESSMENT PHILOSOPHY AND APPROACH 7

A Guiding Philosophy for Assessment in the Foundation 7

Assessment at the Level of Initiatives and Programs 9

Assessment at the Project Level 12

Procedures and Mechanisms 13

III MOTIVATING AND SUPPORTING ASSESSMENT 17

Roles and Locus of Control 17

Incentives and Rewards 19

Resources 21

IV MEETING THE CHALLENGE 25

Prospects for Improvement 25

Benefits of Improving the Assessment of Science Education Initiatives 26

CONTENTS (Continued)

PART TWO: DESIGN CONSIDERATIONS 29

V ASSESSMENT QUESTIONS 31

Three Perspectives on the Assessment Target 31

The Operation and Overall Effects of Science Education Initiatives 33

Zooming Out: The Area of Investment to Which the Initiative Relates 37

Zooming In: Close-ups of Individual or Institutional Change and the Initiative's Operation 38

How the Three Perspectives Can Be Used to Document and Examine Initiatives 41

VI PROCEDURES AND MECHANISMS 45

Comprehensive Assessment Studies 46

Documentation Activities 51

Short-Term Focused Assessment Activities 54

PART THREE: A PILOT TEST OF SHORT-TERM FOCUSED ASSESSMENT PROCEDURES 63

VII LIMITED CASE STUDIES 67

Case Studies of Initiatives in Midstream 67

Assessing Support for Collaborative Exhibit Development: The ERC Case 68

Lessons Learned for Further Applications of Limited Case Studies 72

CONTENTS (Concluded)

VIII EXPERT ANALYSES AND SYNTHESES 75

Describing the Domain of Investment Through Synthesis and Analysis of Secondary Data: A "Macro View" of Informal Science Education 75

A Market Analysis of a New Investment Area: Examining the Potential of Videocassette Technology as a Vehicle for Home Science Learning 81

Reflections on the Further Use of Expert Analyses 84

IX WORKING SEMINARS 87

A Cross-Program Principal Investigators' Meeting: Examining Support for Projects That Establish Linkages Between Schools and Informal Educational Institutions 87

An Expert Mini-Conference: Approaches to Assessing the Effects of Informal Science Education on Individual Learners 93

Reflections on Further Use of Working Seminars 98

References 101

Appendix: Acknowledgments. 105

LIST OF TABLES

Table I-1. Illustrative Assessment Questions Concerning Initiatives in Science Education 4

Table II-1. Summary of Recommendations for Improving NSF's Assessment Philosophy and Approach 8

Table III-1. Summary of Recommendations Regarding Ways to Motivate and Support Assessment in the Foundation . . . 18

Table III-2. Three Options for Funding of Assessment Within the Directorate for Science and Engineering Education . . . 23

Table V-1. Generic Assessment Topics 34

Table VI-1. Short-Term Focused Assessment Procedures 55

Table VII-1. Practical Considerations in Conducting Limited Case Studies, Based on Pilot Test Example 74

Table VIII-1. Practical Considerations in Supporting Expert Analyses, Based on Pilot Test Examples 86

Table IX-1. Practical Considerations in Conducting Working Seminars, Based on Pilot Test Examples 99

LIST OF FIGURES

Figure V-1. An Analytical "Zoom Lens": Three Perspectives on NSF's Science Education Initiatives 32

Figure V-2. Model of the Operation and Overall Effects of an NSF-Supported Initiative in Science Education 35

Figure V-3. Model of Individual Learning and Change Implied by NSF's Initiatives in Informal Science Education 40

PART ONE

AN APPROACH TO ASSESSING SCIENCE EDUCATION INITIATIVES IN THE NATIONAL SCIENCE FOUNDATION*

This report presents recommendations to the National Science Foundation (NSF) to guide it in assessing its initiatives in science education.**

The report outlines appropriate goals, procedures, arrangements, and resources necessary to establish an effective set of assessment practices that (1) build on existing assessment activities in the Foundation, (2) fit with agency culture and constraints, and (3) are both comprehensive and practical.

Scope of the Report

We use the term "initiative" loosely to describe all forms of support for education, including targeted funding for a particular problem, such as the preparation of middle school science teachers, and support for less focused activities, such as graduate fellowships, innovative materials development, or research experiences for undergraduates.

Unlike our earlier analysis of investment opportunities in K-12 science education (Knapp et al., 1987a, b, c), our ideas about improving assessment apply to initiatives at any level of education from elementary grades through postgraduate study. Our recommendations can be used by any directorates within the Foundation that make such investments.

Our task did not include the assessment of other activities supported by the Foundation--basic scientific research, the establishment of research centers, etc. To an extent, these investments call for different forms of assessment. Nonetheless, the ideas presented in this report may be used to improve assessment of these activities as well. As some Foundation planners have already recognized, funding for scientific research raises the same basic questions of payoff to investment that are often reserved for initiatives in education. Investments in the production of scientific knowledge, interinstitutional collaboration, and other forms of support for science can parallel the complexity of educational initiatives. In such instances, the approach and procedures we outline in this report have great utility.

---

* *Part One* also appears as a separately bound *Summary Report*.

** In this report we use the terms "science education" and "education in the sciences" to include education in mathematics, the natural sciences, engineering, and technology.

What We Mean by "Assessment"

By "assessment" we mean the following:

> **Any systematic effort to gather, interpret, and report information or evidence intended primarily to contribute to decisionmaking about the Foundation's programmatic support.**

Our definition thus includes a broad range of activities, from short-term, low-cost activities such as syntheses of expert opinion to large-scale contracted evaluation studies. Activities carried out by NSF staff or third-party grantees and contractors are included in the scope of our definition.

We do not, however, equate assessment with all forms of NSF-funded "research" or "studies" in science education, although there is clearly overlap. For example, studies of the status of science education nationwide, often reported in *Science Indicators* (e.g., National Science Board, 1985), are not intended primarily to inform the Foundation's decisionmaking, yet they contribute a great deal to understanding the context surrounding NSF's support for science education.

We also do not restrict assessment of science education initiatives to quantitative studies that take student achievement as the primary outcome of NSF grant support, although these studies provide a useful perspective on certain investments. Rather, we emphasize assessment approaches that assemble quantitative and qualitative information from a variety of sources.

Organization of the Report

In this part of the report we present our argument for improving the Foundation's approach to the assessment of science education initiatives. All of our recommendations to the Foundation appear in these four sections.

Part Two details two sets of design considerations. The first discusses the framing of assessment questions from different perspectives, and the second describes design options and elaborates the strengths and weaknesses of different procedures and mechanisms for carrying out assessments.

Part Three reviews the results of SRI's pilot test of short-term, focused assessment procedures in informal science education. We describe the procedures we used, illustrate the findings, and draw methodological lessons for further application of these procedures.

I THE SPECIAL CHALLENGE OF ASSESSING INITIATIVES IN SCIENCE EDUCATION

The assessment of science education investments presents the Foundation with a challenge unlike the task of assessing support for basic scientific research. Assessments must answer different kinds of questions and therefore be designed with the unique characteristics of this investment area in mind.

Funding for science education is meant ultimately to change the educational experiences and outcomes of learners. NSF seeks to accomplish this goal by investing in the development of curricula, the continuing education of faculty members or school teachers, the production of science television shows or museum exhibits, and opportunities for the enrichment of promising students from middle school through the postgraduate level. The connections between investments and results, however, are often subtle and not easy to see.

Audiences inside and outside NSF raise interesting and difficult questions about the connections between NSF initiatives and these outcomes, such as those listed in Table I-1. As the questions in the table illustrate, audiences want to know more than the amount of growth in the scientific talent pool that can be attributed to NSF's funding. Some questions concern the likelihood that individuals will learn something or change their behavior as a result of NSF-supported activities. Other questions ask about grantees' implementation of NSF-supported activities or an initiative's overall impacts on educational institutions. Still others seek to understand how NSF's initiatives are related to a larger domain of activity. These audiences ask "What happened?" and "How?" or "Why?" as often as they ask questions conventionally associated with assessment, "Does it work?" and "What is the ultimate payoff?"

The most appropriate approach to answering these questions varies. In many instances, good counts of activities or individual participants are sufficient. But often, the question calls for an intensive examination of the way an activity is carried out and the way participants respond to it.

Forces for Improvement in the Foundation's Assessment Practices

Two sets of forces are pushing the Foundation toward a more thoughtful and comprehensive use of assessment to guide initiatives in science education. The first is internal: as we noted in our report on the Foundation's K-12 investment options (Knapp et al., 1987a), NSF has begun to act more strategically in its support for science education and in so doing has a greater chance of significantly improving science education nationwide. **Part of being strategic is knowing whether and how the strategy holds up. Assessment of investments in midstream and at their conclusion is thus a natural and integral element in the Foundation's attempts to act strategically.**

Table I-1

ILLUSTRATIVE ASSESSMENT QUESTIONS CONCERNING INITIATIVES IN SCIENCE EDUCATION

Postgraduate Level	
▪ Postdoctoral research fellowships	▪ Are current stipend levels a sufficient motivator to attract the best minority graduates into scientific work? Is the fellowship mechanism equally effective in all disciplinary areas? Why or why not?
Undergraduate Level	
▪ Development of curricula (e.g., calculus)	▪ How readily are new undergraduate curricula picked up across institutions or adapted by them? What are the most effective ways of encouraging the spread of these curricula?
▪ Research experiences for undergraduates	▪ What types of undergraduates participate in NSF-supported research experiences? How do these experiences alter students' further educational choices?
K-12 Level	
▪ Development of elementary curricula through partnerships with publishers	▪ How do developers and publishers interact in publisher partnerships? What tradeoffs occur under this arrangement between innovative development and widespread distribution of new curricula?
▪ Training for leadership teachers	▪ To what extent do leadership teachers have a "multiplier effect" in outreach to their colleagues on returning to their schools? What factors help or hinder their efforts?
▪ Production of science television series for children	▪ What do children take away from viewing science television?
▪ Science enrichment experiences for bright high school and junior high school students	▪ In what ways do intensive science enrichment experiences affect decisions about scientific careers? Does a greater proportion of these students pursue scientific majors in college than of others who do not participate in enrichment projects?

The Foundation's scientific "culture" and the staff's professional concern to examine and understand the rationale behind NSF's investments provide another internal force for improved assessment. Foundation staff tend to prize professional competence over bureaucratic position; **managers at all levels ask themselves hard questions about the activities they support, which cannot be answered at the proposal review stage. These managers want and need good assessment to answer their questions** (in fact, several of the questions in Table I-1 were originally posed by Foundation planners and program managers).

The second set of forces for improved assessment is external: agencies and arms of the government on which the Foundation depends for its resources--Congress and the Office of Management and Budget, in particular--want to know what initiatives in science education are accomplishing (e.g., General Accounting Office, 1984; House Appropriations Committee, 1987; Senate Appropriations Committee, 1987). **Whether or not specific questions are asked, these bodies need to be convinced in the annual budget process that investments in science education are sound. Good assessment can play a central role in both rationale-building and reporting to these audiences.**

Audiences in the relevant professional communities--curriculum developers, disciplinary scientists engaged in education, teacher educators, publishers of science tests, for example--also ask important questions about NSF's investments in science education. **The Foundation exerts "intellectual leverage" over those professional communities in proportion to the depth and breadth of the publicly available knowledge about what it supports.** Because the "professional community" in science education is so large and diverse, existing professional networks cannot be counted on to spread the word, much less to determine accurately what NSF initiatives have accomplished. This fact redoubles the need for systematic and effective assessment in this area of NSF support.

Groundwork for Improving Assessment of Science Education Initiatives

Recent developments in the Foundation lay the groundwork and provide some models for more comprehensive improvements in the Foundation's assessment approach. Consider, for example:

- *The creation of an Office of Studies and Program Assessment (OSPA) in the Directorate for Science and Engineering Education (SEE).* SEE has created an office with a budget of its own, the responsibilities of which include the sponsorship of assessments, the gathering and analysis of information in-house, and the provision of technical support to other SEE staff.

- *Assessment activities in SEE.* With the help of OSPA, contracted assessment studies have been initiated to examine the operation of the College Science Instrumentation Program and Presidential Young Investigators' Awards. Other assessment activities have included a few grant-supported studies and several commissioned papers on planning-related topics, not to mention the two-part SRI study.

These developments parallel activities elsewhere in the Foundation that will help to build NSF's capacity for assessing science education initiatives. For example, an ambitious restructuring of the Foundation's management information systems (MIS) capability is currently under way, which will enable program managers throughout the Foundation to assemble prompt and accurate descriptive information about the projects they are supporting. Although evaluation of the Foundation's scientific investments tends to lag behind assessment in education, there are even some promising experiments with assessment of NSF's scientific research initiatives, such as the Industry-University Collaborative Research Centers (IUCRC). Each of these centers currently employs a part-time evaluator to document the center's progress. The evaluators meet periodically to share findings and develop cumulative understanding about the initiative. These kinds of activities have come about with the full support of the Foundation's leadership.

What the Foundation has accomplished so far provides some models and the starting points for developing a more comprehensive set of assessment practices for science education, but the process of developing these practices is far from complete. As detailed in the following two sections, a series of additions and adjustments to current assessment practices and policies would put in place the rationale, tools, and organizational arrangements to carry out effective assessment over the long term.

II ASSESSMENT PHILOSOPHY AND APPROACH

Effective assessment of science education initiatives in NSF begins with a clear philosophy about the relationship of this function to programmatic grantmaking. On the basis of this philosophy, one can suggest appropriate approaches to assessment at the level of initiatives or programs and also at the level of individual projects supported under these initiatives. Finally, these approaches, in turn, imply particular procedures and mechanisms.

We summarize our recommendations about assessment philosophy and approach in Table II-1, then briefly explain each one below.

A Guiding Philosophy for Assessment in the Foundation

Because "assessment" means many things to different people, it is easy to be unclear about the purposes for this activity and approaches to it. We propose a guiding philosophy that views assessment as follows:

- *Assessment is an integral part of proactive, strategic support for science education.* This means that assessment is a process of learning about what NSF supports, in order to clarify its strategy and influence decisions about future areas of investment. As such, it is as central to what the Foundation does as the grantmaking process itself.

- *The Foundation should design and use assessments to inform future action--in particular, program planning, resource allocation, reporting, and program justification.* To accomplish these purposes, NSF must frame assessment questions to anticipate future action issues, design assessment to fit the timetable of decisionmaking, and establish routines that encourage the availability of assessment information to those who may desire it.

- *Assessments should emphasize learning from initiatives rather than making summary judgments about them (even though what is learned will naturally contribute to the judgment process).* When assessment falls into a judgmental mode (which can easily happen), individuals feel threatened and a great amount of energy is expended countering or subverting the implied attack. It is preferable to aim for description and explanation--what happens (or is likely to happen) and why.

Table II-1

SUMMARY OF RECOMMENDATIONS FOR IMPROVING NSF'S ASSESSMENT PHILOSOPHY AND APPROACH

Guiding Philosophy

(1) Assessment is an integral part of proactive programmatic grantmaking.

(2) Assessment should be future-oriented and be designed to facilitate planning, resource allocation, program justification, and reporting.

(3) Assessment should emphasize learning about initiatives rather than making judgments about them (although what is learned may contribute to these judgments).

(4) Assessment should assemble, from a variety of sources, a "mosaic of evidence" about its initiatives.

Assessment at the Initiative and Program Levels

(1) Focus on logically related investments within and across grant programs.

(2) Document initiatives by developing a basic set of quantitative and qualitative information about what is supported.

(3) Examine the logic, rationale, and assumptions underlying initiatives.

(4) Study selected projects in depth to exemplify an initiative's accomplishments or examine its assumptions.

Assessment at the Project Level

(1) Decrease the reliance on principal investigators as the basic source of assessment information.

(2) Focus project-based assessments on improving the project itself by encouraging "formative evaluation" of some kind.

(3) Make it possible for principal investigators to furnish NSF with standardized descriptive information about their projects.

Procedures and Mechanisms

(1) Assemble evidence from a combination of (1) comprehensive assessment studies, (2) documentation activities, and (3) short-term, special-focus activities.

(2) Establish mechanisms to carry out all three of these on an ongoing basis.

- *The Foundation should assemble, from a variety of sources, a "mosaic of evidence" about the initiatives it undertakes rather than relying on a single source of evaluative information.* NSF's science education initiatives are too complex to submit to easy answers derived from a single source or study. For example, although it is possible to study leadership teacher training through a single comprehensive study, the Foundation can gather evidence about this initiative more efficiently and promptly through a combination of separate assessment efforts that examine different aspects of this initiative simultaneously.

When this philosophy is translated into operational terms, it means different things at the level of initiatives or programs and at the level of individual projects funded under these initiatives.

Assessment at the Level of Initiatives and Programs

The Foundation should increasingly aim assessments at identifiable initiatives and, in some instances, at grant programs taken as a whole. The Foundation is supporting some studies at this level, such as those undertaken by SEE (noted in the preceding section), but a more varied and comprehensive effort to document and examine initiatives needs to be in place if the kinds of questions posed earlier are to be answered as a matter of course.

Focus on Logically Related Investments Within and Across Grant Programs

NSF's assessments are most likely to inform future strategic decisions if they focus on the logically related investments that compose the Foundation's strategy. This may mean examining formally declared initiatives--as in the case of the special solicitations issued by SEE to address elementary science materials development or middle school teacher preparation--or sets of projects that happen to tackle the same area, as in the case of teacher enhancement projects that train elementary mathematics teachers.

Under some circumstances, the grant program is the logical unit for assessment. SEE's College Science Instrumentation Program, for example, issues one kind of award to a large number of postsecondary institutions with a single goal in mind: upgrading the instructional instrumentation used in college laboratories. But more often, examining the program as a whole lumps together unlike types of investments and also makes it difficult to see the connections between programs.* SEE's Instructional Materials Development Program, for example, supports large-scale curriculum development efforts through publisher partnerships (in response to a particular solicitation), as well as the development of innovative instructional materials by individual principal investigators or small project teams. In such instances, assessments ought to examine the two approaches to curriculum improvement separately.

---

* A mechanism exists--the program oversight committee review--to examine the operations of programs taken as an administrative unit. Although this procedure cannot carry out assessments in great depth, it can be and has been used to address important prospective assessment questions.

Develop Descriptive Documentation of Initiatives

If it does nothing else, NSF needs to develop a basic descriptive data base on what it supports. Under each initiative, the Foundation should document, first of all, the "basic facts" about project activities that are easily counted--for example, numbers of participants in teacher enhancement workshops, the proportion of young scholars who are from minority backgrounds, or the amount of matching funds put forth by colleges receiving instrumentation improvement grants. Standardization of terminology is vital to make simple counting meaningful across projects, and to avoid inadvertently duplicated counts of participants who repeat in any program.

But just as important are the qualitative characteristics of the activities NSF supports. For example, the Foundation should try to learn what types of follow-up the organizers of teacher enhancement workshops engage in, the nature of young scholars' research (or other enrichment) experience, and the ways new instrumentation is used in college laboratories.

This kind of information has rarely been gathered in the past and would be especially useful to NSF. For example, the Foundation found itself in the position in the early 1980s of being unable to report to Congress even such basic statistics as the number of teachers who participated in summer institutes during the 1970s (General Accounting Office, 1984). Some of the assessment questions listed earlier in this report ask for similar information. How many and what types of undergraduates participate in NSF-supported research experience programs? How many teachers are reached by NSF-supported leadership teachers after they complete their training? If answers to these questions are routinely available, NSF can not only meet a number of its reporting and planning needs, but also establish a baseline to be included in more complex assessment studies.

Examine the Logic, Rationale, and Assumptions Underlying Initiatives

Rather than study the effects of each project funded under a certain initiative, NSF is typically better off studying the logic, rationale, and assumptions on which the initiative rests. The basic questions are these: Is the initiative sound? How and why does it work the way it does? What lessons can be learned from it for improving it and other related investment thrusts?

This approach means looking at initiatives from several perspectives at once. To take a brief example from the list of assessment questions in Table I-1, NSF's initiative to develop new undergraduate calculus curricula can be looked at on several levels. NSF can study the operational logic of this initiative to determine whether the right proposers are likely to respond to NSF program announcements, whether exciting curricula will be developed, and, if so, whether these will get published or otherwise disseminated. One can also examine assumptions about the need or demand for new calculus approaches. At the same time, the initiative rests on other assumptions about the way new curricula are adopted or adapted at the undergraduate level, and even about the way undergraduates view the learning of calculus. Effective assessment of this initiative means examining all these assumptions to the extent possible. If one key assumption doesn't hold--for example, if the demand isn't there, even though good developers are interested and appropriate distribution mechanisms exist--then the soundness of the initiative (in its current form) can be questioned.

There are various advantages to aiming assessments at this target. First, and most important, it leads the Foundation to consider the reasonableness of its investment strategies, without becoming immersed in the details of all the projects that carry out those strategies. Second, the focus on underlying logic and assumptions allows assessment to be done more efficiently, for example, by gathering data on a few key projects and by looking simultaneously at other sources of information (see discussion of procedures and mechanisms below). Some key assumptions can be tested by examining projects that have no NSF support at all (see the fifth question in Table I-1). Thus, the Foundation need not wait until all the projects are completed under a given initiative before it is able to develop evidence on which further planning or resource allocation can be based.

Study Selected Projects in Depth

Under certain circumstances, the Foundation may want to study an initiative by examining the activities and results of particular projects in great detail. Such examinations are especially useful when a project constitutes a critical "test" or demonstration of the model underlying an initiative. An example of a recent assessment undertaken by SEE illustrates this approach:

- A project grant (Crane, 1987) supported a recent exploratory study of the science television series "3-2-1 Contact!" and its effects on young viewers. Although not explicitly evaluative, this study documented in great detail many aspects of NSF's investments in science broadcasting.

This is only one instance in which a project comprises a "critical case" deserving careful assessment; others come readily to mind, such as some of the leadership teacher training projects the Foundation has supported over the past 5 years. Rather than study such projects on an occasional basis, this kind of assessment could be done more frequently and systematically to develop in-depth information about the operation of an initiative in the field.

Assessment at the Project Level

The emphasis we place on assessment at the initiative level changes the approach to assessing individual projects. Currently, NSF relies too heavily on self-assessments done by each project. In SEE, if not elsewhere in NSF, most principal investigators are required to conduct a self-evaluation of their projects, which they submit as part of the project's final report.

For various reasons, project-level self-assessments are not a useful way to answer most questions about the Foundation's support for science education. NSF should therefore change its approach to project-based assessments. For one thing, although self-assessments carried out by each principal investigator can provide useful insights, they are unlikely to yield a "big picture" view that the Foundation needs to understand the effects of its initiatives.

Decrease Reliance on Project-Based Self-Assessments

Self-assessment by NSF grantees tends to fail because of a basic fact of life: principal investigators typically have neither the technical skills nor the motivation to conduct a thorough evaluation of their own work. It would be costly and difficult to provide enough resources and technical assistance to all principal investigators to improve their assessment activities (even if they wanted to). But even if most principal investigators or their project teams could be made into capable evaluators, their efforts might not, in the aggregate, lead to better understanding of NSF initiatives. For example, one does not necessarily get the best answers to questions about NSF's support for science teacher networks by asking network creators to critique their own efforts (even though any reasonable assessment would consider their views as *one* perspective on networks' efficacy). Not only do they lack a degree of objectivity with regard to their own work, they lack the larger perspective of a funds-granting agency, which must take many things into account as it weighs the value of its investments or considers how to improve them. Even more important, one does not need a report from *all* network projects to learn whether the logic or assumptions underlying this type of initiative are sound.

Encourage Projects To Do Formative Evaluation for Their Own Use

Nonetheless, project self-assessments can contribute to a more modest goal: helping the project team reflect on what they are doing and make mid-course corrections. The value of this kind of "formative" assessment has been effectively demonstrated in some projects funded by SEE to develop curricular materials, science television shows, and museum exhibits. In such instances, assessment information is tailored to the specific needs and circumstances of each project.

The example set by these projects could be followed more widely by NSF-supported projects in science education, especially if the Foundation encouraged this kind of evaluation as a legitimate use of project funds. (Principal investigators who lack assessment expertise would still need to seek assistance for this activity.) Formative evaluation to serve project purposes need not be elaborate and costly; a variety of useful techniques exist that can help project staff do a thoughtful, reflective job (see discussions in Volume 2, Sections I and V).

Enable Projects To Furnish NSF with Basic Descriptive Information

To document what it supports, the Foundation needs some descriptive information on all projects. For obvious reasons, it is difficult to aggregate information about each project when assessment designs are developed locally to suit the project's particular characteristics. A promising alternative exists: NSF can encourage project directors to supply the Foundation with standardized descriptive information about project activities, participants, resources, impacts, etc., in response to data requests from the Foundation (or a third party acting in a documentation role). The Foundation could make it easy for project directors to furnish this information by developing standardized forms, by supporting telecommunication links, and by other devices (see below).

Procedures and Mechanisms

The approach to assessment we have outlined requires a flexible array of procedures and mechanisms. To assemble a "mosaic of evidence" about its science education initiatives, the Foundation will need more than the few contracted studies now in place. We recommend that NSF carry out assessments through a combination of comprehensive assessment studies, documentation activities, and short-term focused analyses. A detailed discussion of these three appears in Section VI; we briefly review the categories below.

The first of the three--comprehensive assessment studies carried out through grants or contracts--has clear precedents within the Foundation and requires little further explanation. The advantages of this approach to assessment are obvious: it provides the most complete and credible data about initiatives and it is highly visible. At the same time, there is a long time between procurement and final results. In addition, the RFP mechanism, by which most such studies are supported, is cumbersome and relatively inflexible. As a consequence, comprehensive studies should never be thought of as the only--or even the primary--way by which the Foundation's assessment questions can be answered.

Documentation activities complement comprehensive studies by generating an ongoing descriptive record of the activities NSF supports. Three sources of this information seem especially promising, and should be considered as NSF plans its approach to assessment:

- *Improved MIS capabilities.* Already under way, improvements in MIS capabilities can be used to tally, track, compare, and report on the characteristics of grantees and other kinds of information received as part of the proposal process.

- *Documentation grants.* Grants (or contracts) to third-party researchers can be used to assemble particularly detailed or qualitative types of documentation, such as accounts of the collaboration between publishers and developers in partnership arrangements.

- *Data collection systems.* For certain kinds of initiatives, e.g., those involving services to individual teachers or students, ongoing data collection systems can help to track cohorts of participants and gather other kinds of descriptive information about projects.

The Foundation does not yet use any of these devices to document support for science education, although several have been considered and steps have been taken to improve the Foundation's MIS (though not with the assessment of science education in mind). Documentation activities are not difficult or especially costly to set up, and would provide a basis for further, more focused assessment work over the long term.

The third category of activity--short-term focused assessments--complement comprehensive studies in a different way. These activities can be done in a matter of months, by one or a few individuals. Four types of activities within this category have wide application to the assessment of support for science education:

- *Limited case studies.* Brief site visits to selected samples of projects (e.g., all of which aim at a common target) or case reviews of key projects or institutions can shed light on the implementation of NSF-funded activities, individual learning, and interaction between participants and NSF-supported resources.

- *Quick-response surveys.* Either by phone (for smaller samples of projects and individuals) or by mail (for larger samples), simple surveys can answer questions about project accomplishments or the experiences of individuals who participate in these projects.

- *Expert analyses and syntheses.* Many assessment questions can be answered by expert judgment and analysis of information from existing data sources: for example, statistical analyses to generate a profile of the areas in which NSF invests its resources, literature syntheses, meta-analyses of research results, and market analyses.

- *Working seminars.* Groups of experts meeting for short periods of time can address questions that require group interaction and discussion: for example, meetings of principal investigators from thematically related projects or mini-conferences of experts related to a particular assessment topic.

Although, in principle, NSF staff can carry out these procedures themselves, NSF is better off using other means--in particular, the following three mechanisms: (1) **adjunct staff** (who come to the Foundation for short periods of time to conduct analyses or seminars); (2) **task ordering agreements** (that secure a third-party organization to do small tasks as needed); or (3) **personal services contracts** (which compensate an individual for a particular limited task). The Foundation has made use of all three on occasion, but seldom with assessment of science education activities in mind.* By drawing on its own experience and that of other agencies, the Foundation could put these mechanisms in place readily.

---

* An exception is SEE's use of personal services contracts to support analyses for *Science Indicators* and to support commissioned papers on long-range planning issues.

III MOTIVATING AND SUPPORTING ASSESSMENT

To improve science education assessment in the Foundation, the right combination of expectations, incentives, and resources must be in place. Otherwise there are natural and understandable tendencies for this function to be viewed as something extra, something to be feared, or a drain on valuable resources.

However, **if people understand the roles they are expected to play in assessment, see rewards for carrying out these roles, and receive adequate funding and technical advice, then a "climate of support" for assessment will develop.** Generally speaking, the current climate in the Foundation is not as supportive of assessment in science education as it could be, but such a climate can be cultivated. When that happens, assessment will become an integral part of the Foundation's efforts to improve science education.

We present below our recommendations regarding staff roles and locus of control, incentives and rewards, and resources. For easy reference, the recommendations are summarized in Table III-1.

Roles and Locus of Control

If assessment is to become part of NSF routine, this activity must be collaborative, and at the same time staff at various levels must play somewhat different and independent roles. Individuals at one level in the Foundation know only part of the "story" about any particular initiative. At the directorate level, for example, planners and managers typically understand the "politics" of a given initiative and its place in overall investment plans, but not its details--what types of groups are funded, what these groups are undertaking, etc. These details are the province of program officers, who may not have as good an overview of the initiative in relation to other aspects of NSF's overall strategy in science education. At each level, individuals are likely to pose important questions that are not raised at other levels nor are necessarily relevant there.

Assessment must be collaborative yet differentiated for another reason. No one wants to feel like the passive subject of scrutiny by others, especially by superiors in the Foundation's chain of command. Individuals are more willing to cooperate with assessment activities when they themselves contribute to these activities.

Table III-1

SUMMARY OF RECOMMENDATIONS REGARDING WAYS TO MOTIVATE AND SUPPORT ASSESSMENT IN THE FOUNDATION

Roles and Locus of Control

(1) Expect every professional to contribute to assessment, at least in setting agendas for assessment and in interpreting results.

(2) Encourage each level in the Foundation to initiate assessment activities that answer questions relevant to that organizational level.

(3) Make a sufficient number and range of specialists available to provide technical support to those who need it.

Incentives and Rewards

(1) Adjust or, if necessary, restructure managerial and staff assignments and workload to make assessment activities an essential part of the grantmaking process.

(2) Reward individuals and organizational units in the Foundation for carrying out and using assessments effectively.

Resources

(1) Allocate adequate resources to assessment--in the range of 2% to 5% of total funds spent for science education support.

(2) Disperse the resources for assessment among the budgets for specialized units (e.g., in SEE's Office of Studies and Program Assessment), program budgets, and discretionary accounts available to divisional or directorate-level managers.

Because of these facts, **assessment will be effective and sustained in the Foundation only if it is the joint result of actions by many individuals at various levels in the organization rather than the sole responsibility of a few specialists.** Practically speaking, this means that the Foundation should:

(1) *Expect everyone to contribute to assessment.* All program staff and managers would be expected to participate in assessment--at a minimum, by contributing to the development of an assessment agenda and to the interpretation of assessment results that pertain to their sphere of activity.

(2) *Encourage each level in the agency to initiate assessment activities that answer questions relevant to that organizational level.* Individuals at each level would be empowered (through appropriate resources and incentives, as discussed later in this section) to initiate and conduct assessment activities that serve their immediate needs. Within each program and division (or office), staff would be strongly encouraged to undertake one or more such activities each year.

(3) *Make a sufficient range and number of specialists in assessment available to provide technical support to those who need it.* Specialists with particular expertise in assessment (for example, staff of the Office of Studies and Program Assessment in SEE) would be expected to provide technical advice and ongoing assistance to others (as OSPA now does), and in some instances to coordinate assessment efforts. Such individuals would devote a majority of their time to sponsoring and conducting assessments, or helping others to do so.

A system of dispersed control over assessment is not without drawbacks or tensions. We recognize that this kind of activity always has the potential to become involved in issues of organizational competition and control. However, if assessment activities are, in fact, initiated by staff at different levels, then the danger of centralized or "top-down" control over assessment is avoided. If staff are routinely invited to help set assessment agendas and also to interpret results, then this function will lose some of its threat. If staff at all levels have resources with which to undertake assessments that serve their own needs best, then they exercise effective control over at least some of the assessments that are done.

Incentives and Rewards

Clarifying everyone's role and the locus of control in the assessment function provides one set of incentives for contributing to this activity: people are more likely to participate if it is part of their job description and if they exercise some control over it. But another natural disincentive has a crippling effect on any attempt to carry out effective assessment in the Foundation: insufficient time to undertake assessment activities.

NSF's professional staff engaged in support for science education are a hard-working group; the complexity of the proposals they receive requires a great investment of staff time. Most of them believe, with some justification, that there is not much time for anything more in their workdays, including assessment. Those who care most about assessment try to find time for it, but typically their days are consumed by the demands of processing proposals and other staff or management tasks. The squeeze on professional time is exacerbated by other things, such as the fact that the Foundation's funding for science education has been growing rapidly. This growth means that staff now in place may have to process more proposals, before new staff can be brought on to handle the increased load.

Realistically, time for assessment will be found only if managers make time for this function. That will happen only if NSF indeed adopts a more proactive, strategic model of grantmaking. To overdraw the contrast (for sake of explanation), NSF need not set aside time for assessment if it makes grants in a largely "reactive" fashion, that is, by funding good people with interesting ideas and trusting that they will contribute to the improvement of science education. Under this model, assessment is, in fact, an extra. If, on the other hand, NSF assumes a more proactive funding posture (and it has begun to do so in many aspects of its science education support), then assessment is an inescapable part of program managers' jobs. Not only must they make grants, but they must also check to see whether their initiatives are sensible, appropriately targeted, and accomplishing (or likely to accomplish) something useful. Furthermore, they must develop information that would help to plan the next initiative on the drawing board.

At present, program staff in science education appear to be in transition between the two conceptions of their job. Although they tend to spend their time more in accordance with the reactive model described above, many engage in proactive grantmaking activities as well. If the transition continues (and we urge it to), the process will be gradual, and the limitations on time for the assessment function are likely to be felt in some form for some time to come. The Foundation can take two kinds of steps to facilitate the transition:

(1) *Adjust or, if necessary, restructure staff assignments and workload to make assessment activities an essential part of the grantmaking process.* Because doing this kind of restructuring involves basic questions of staff time allocation among all functions, it lies beyond the scope of our study to suggest what adjustments or restructuring might be appropriate. But various possibilities come readily to mind--for example, assigning certain individuals in each programmatic division a large role in assessment and correspondingly fewer responsibilities for other activities.

(2) *Reward individuals and organizational units in the Foundation for carrying out and using assessments effectively.* It is conceivable that individuals could be rewarded for competent assessment in much the same way that they are now recognized for their skill in making grants. Organizational incentives (including funding incentives) can also be created for

developing and using good assessment information. Once again, the specific form for rewards and incentives can be worked out only as part of the overall reward system that operates throughout the Foundation, as well as within individual directorates.

Resources

Finally, sufficient funds must be allocated to assessment activities. How much does it take to support and sustain an effective assessment function? We believe that NSF can and should spend a higher percentage of its annual budget than it now does for assessment activities, regardless of the total budget level. Similarly, it seems appropriate for NSF to gradually increase the number of assessment activities that it undertakes (including relatively low-cost special-focus activities).

Our general answer to the question of how much to allocate is this: **effective assessment practices will require between 2% and 5% of the total funding for science education support.** These funds can come partially from program budgets (e.g., where program staff support assessment activities through grants or add-ons), from divisional or directorate-wide discretionary funds (e.g., for assessment contracts, task ordering agreements, personal services contracts), and from specialized accounts (as in SEE's Office of Studies and Program Assessment). As we argued above, the funding should not be centralized, although for obvious reasons the activities of designated specialists or offices might account for the bulk of assessment funding.

Our recommendation that NSF increase the proportion of its science education budget devoted to assessment is made without regard to the overall level of funding available for science education. At any level, assessment is a "core function" that is critical to effective investment of the Foundation's resources.

To illustrate how NSF might address the question of resources, we lay out options that might be considered by SEE, the directorate that controls the largest share of the Foundation's resources for science education. The Directorate can invest in assessment at several levels. To estimate each level, we distinguish several types of assessment activity: (1) large studies of entire initiatives or programs, costing $250,000 or more per year (often for several years); (2) medium-size study contracts (or grants) in the range of $100,000 to $250,000 per year, which may focus on smaller clusters of projects, very large individual projects, studies of an entire domain of investment (e.g., teacher preparation, informal science education), or other assessment topics; (3) data collection system projects, the costs of which are likely to be in the same range as those of medium-size studies; and (4) short-term focused activities, costing less than $100,000 each, including meetings, visits to exemplary projects, case studies, small-scale surveys, commissioned papers by experts, etc. (As discussed earlier, activities in the last category may be administered through a single task ordering agreement, but can still be budgeted and considered independently.)

Three options for funding assessment in SEE are summarized in Table III-2 below. The options vary in terms of the level of resources and range of assessment activities across programs and divisions within the directorate. The first option enables very little of what we have proposed to be accomplished. The two higher levels of investment in assessment, on the other hand, come closer by stages to the degree of support implied by an assessment function of the sort we have described.

- *Minimal funding.* Under this option (which comes closest to SEE's current allocation to the assessment function*), SEE could support relatively little assessment activity. Completion of one medium-size study each year, one large study every 3 years, and a few special-focus assessment tasks would cost about $1.1 million annually. At this rate, it would take 12 years for each of the four divisions to commission and complete one large assessment activity.

- *Low funding.* At a budget level of $2.0 million annually, SEE could double the number of large assessments, so that each of the four divisions could commission one every 6 years, while commissioning a medium-size assessment every 4 years. Each division could also support three or four small assessment tasks annually.

- *Comprehensive funding.* At this level--approximately $4.6 million annually--each of the four divisions in SEE could commission a large assessment every third year. (If each of these focused on initiatives within one program, it would take about a decade to study every program.) Each division could also commission annually two medium-size and five to six small assessment activities. In addition, the directorate could support ongoing data collection and analysis projects for two or three of its programs.

The figures shown in the table do not include the proportion of project grants reserved for formative evaluation or response to data requests.

In total, assessment thus requires funds commensurate with a small grant program, although, as we have explained, the function cuts across all programs. **Conceived as an integral part of strategic grantmaking, assessment is as worthy of adequate resources as established grant programs.** This statement does not imply that assessment deserves an equal portion of the budgetary pie. Arguably, assessment should always be limited to a relatively small proportion of overall programmatic expenditures, but the current level of investment in assessment, either in SEE or elsewhere in NSF, is clearly too small to make this function productive.

---

* Not including the portion of grantees' project budgets devoted to self-assessments, nor the funds for the Studies and Analysis program, some of which support work that contributes indirectly to assessment goals.

Table III-2

THREE OPTIONS FOR FUNDING OF ASSESSMENT WITHIN THE DIRECTORATE FOR SCIENCE AND ENGINEERING EDUCATION

	Funding Options (Annual Dollars in Millions*)		
	Minimal	**Low**	**Comprehensive**
Large assessment studies	0.4	0.8	1.6
Medium-size assessment projects	0.5	0.7	1.5
Short-term focused activities	0.2	0.5	1.0
Data collection systems	--	--	0.5
Total	1.1	2.0	4.6
(Grants)**	(0.3)	(0.5)	(1.5)
Percentage of SEE's total funding for science education (in FY 88)	0.8%	1.4%	3.3%

---

[*] Not including the portion of grantees' budgets used for conducting formative evaluation or responding to NSF's requests for data.

[**] Annual amount of the total from grant program budgets, which is used for assessment purposes; other resources for assessment would be allocated to OSPA (although not as part of its grant programs) and to divisional and Directorate-wide discretionary accounts.

Does supporting and sustaining an assessment function mean "taking away from" valuable program investments? Yes, in the sense that, ultimately, resources are scarce and any investment precludes another. No, in the sense that program investments have "value" only if they contribute in some identifiable way to improvement of science education. In addition, the value for the professional community as a whole derives in part from making knowledge about these projects available to a wider professional audience. Given the importance NSF places on maximizing the leverage of its investments, such an allocation level would be fully justified.

IV MEETING THE CHALLENGE

To summarize our argument, the Foundation needs to know what it is accomplishing (or likely to accomplish), and why, when it invests funds in science education--or any area of endeavor, for that matter. Otherwise, its funding will be difficult to justify and future investment decisions will rest largely on intuition, personal experience, analysis of proposal logic, and constituency pressures. In science education, where the Foundation has begun to take on a strategic role in attempting to improve the functioning of educational systems, this kind of knowledge is doubly important. Furthermore, the relevant professional communities need to know what NSF sponsorship and interventions accomplish if they are to benefit from the experience gained through NSF-supported projects.

By broad agreement, the mechanisms within the Foundation for building this knowledge are not yet strong enough. Broadly conceived and intelligently executed, assessment has an important role to play in the process of learning from initiatives, and ultimately in the success of the Foundation's investment strategies.

Prospects for Improvement

If the Foundation agrees that assessment should be given higher priority than at present, the means to improve its assessment practices are at hand. Phased in over a period of years, the following changes in practice and policy will put the right set of practices in place:

- A change in the way managers and professional staff define assessment, its most appropriate targets, and their own roles in it.
- Steps to encourage participation in assessment activity by managers and staff at all organizational levels.
- Adequate access to technical expertise so that managers and staff can get help with assessment activities when they need it.
- Explicit statements of assessment policy for the Foundation as a whole and within the directorates that support science education.
- The development of an annual list of high-priority assessment questions and issues within programs, divisions, and directorates.
- Establishment of mechanisms to document initiatives and to undertake short-term focused assessment tasks on an ongoing basis.

- An adequate allocation of resources to assessment, both to specialized assessment units and to divisional and program budgets.

An improved assessment system will not evolve, of course, without a climate of support for this function. But such a climate will develop only over time, as these steps are taken to establish the function on a firm footing.

To take these steps and develop the right climate of support will require active leadership both at the Foundation level and within the relevant directorates. Leaders in NSF can and must set a tone that encourages the use of good assessment information in decisionmaking; otherwise, "business as usual" will prevail.

Benefits of Improving the Assessment of Science Education Initiatives

As they ponder whether and how to improve the assessment of NSF's science education initiatives, Foundation planners and managers should consider the many advantages of success. The most obvious consequences concern the Foundation's relationship to external constituencies:

- *The outside world may impose fewer assessment requirements on the Foundation.* If they do not get assessment evidence from NSF, Congress or others in the federal policy arena may require the Foundation to do assessments that do not make sense or that NSF does not want to do. By improving its assessment practices, NSF is more likely to be able to control the terms of the assessments and may have to undertake few or no studies that are misconceived or unproductive.

- *NSF's resources for science education are less likely to be called into question.* Without credible evidence of the effects of funding for science education, or even adequate documentation of how these funds are used, funding bodies may be reluctant to continue the flow of resources for science education. The past gives ample indication that a lack of evidence of results decreases the confidence of funders. Recent increases in NSF's funding levels for science education represent a vote of confidence in the Foundation's ability to improve science education; an adequate flow of assessment information to funding bodies will help to make the case for continuing this funding.

- *The Foundation would be less open to criticism that it is not managing its resources well.* The absence of effective assessment might be taken as one sign of ineffective management (a perception that led to the congressional mandate for the SRI study in the first place). The management of support for science education has improved considerably since the hiatus in funding for this area 5 years ago. Effective assessments are one way to display the tangible evidence of these improvements.

By improving its assessment of science education initiatives, the Foundation will be in a better position to manage the complex environment of support and criticism that inevitably surrounds government agency programs. To do so, NSF managers must overcome the natural concern that, in a politicized environment, increased information about science education initiatives will do more harm than good. We acknowledge that such concerns are legitimate and deserve to be carefully weighed. If, for example, most of the Foundation's support for science education were ineffectual, then NSF managers might reasonably conclude that assessment would threaten these investments and should be minimized. However, as our review of NSF's funding options in K-12 science education pointed out (Knapp et al., 1987a, b), NSF has much to be proud of in its history of support for science education. Or, if the only audience interested in assessment results were groups and individuals opposed to funding for science education, then, too, NSF managers would be rightfully concerned about the way assessment results might be used in the public arena. But the advocates of NSF's funding for science education are as interested in this information as the opponents (furthermore, the opponents will push their point of view with or without data). In sum, we believe that NSF has more to gain than to fear by developing good assessment data about its support for science education.

The most important consequence of improved assessment will not be manifested in the perceptions or demands of the outside world, but in the effectiveness of the Foundation's strategies for improving science education itself. In supporting science education, it is not enough to find good people, award them funds on the basis of a careful proposal review, and hope for the best. The challenge for NSF is to maximize the educational impact of its limited resources. This means that the Foundation has to find innovative ways to engineer its investments and develop a repertoire of appropriate and credible practices for assessing them. If NSF can successfully integrate planning, management, and evaluation, it will go a long way toward achieving the real potential it has to improve the science education of the nation's young people.

PART TWO

DESIGN CONSIDERATIONS

The approach to assessment described in *Part One* provides the framework for designing assessment activities. We explore in this part of the report considerations that influence how NSF frames assessment questions, chooses procedures, and establishes mechanisms to carry out assessments.

We do not present specific indicators, measures, or research methods that would be used in assessing particular types of investment in the Foundation's science education "portfolio." Cataloguing these things in a comprehensive way would be an exhausting and counterproductive exercise. Virtually the full array of methods and measures in educational research and evaluation could be used, depending on the assessment questions NSF wished to answer. Furthermore, the range of techniques appropriate to an investment such as research on advanced educational technology would differ greatly from what would be used to examine NSF's support for teacher education, graduate fellowships, or science television series. To resolve these technical matters, NSF can and should turn to relevant experts, and there are inexpensive ways to seek this advice when it is necessary. These individuals, in consultation with Foundation staff, should identify the particular measures and techniques that are appropriate to a given assessment problem (our pilot test examples illustrate how this process would happen in informal science education).

The Foundation's principal "design" tasks are to (1) figure out what questions the assessment should address, (2) identify the types of studies or assessments most appropriate to answering these questions, and (3) create the right mechanisms (funding vehicle, staff arrangement) for getting the work done. The three sections in *Part Two* elaborate our thinking about these three tasks.

V ASSESSMENT QUESTIONS

The philosophy and approach outlined in *Part One* lead the Foundation to pose and answer a different set of questions than is conventionally asked by formal program assessments. Our philosophy shapes the questions being asked in the following ways:

- The emphasis on documenting the activities supported by NSF makes the question "What happened?" central to the Foundation's assessment strategy.
- The prospective orientation for designing and conducting assessments adds the question "What have we learned from what happened that informs our next investments?"
- The emphasis on learning from investments rather than rendering summary judgments about them places priority on questions that ask *why* something worked (or didn't), *in what ways* it worked (or didn't), and *what it means* to "work," rather than simply *whether* it worked or not.
- By focusing on the logic, rationale, and assumptions underlying initiatives, NSF asks questions about the soundness of its strategy, rather than confining questions to a narrow accounting of the use and impact of funds.

In this section we outline a framework for generating such questions, along with examples applied to NSF's investments in K-12 science education. We emphasize the importance of posing questions not only about the initiative itself, but also about the area of science education to which the initiative relates as well as the mechanisms for change implied by the initiative.

Three Perspectives on the Assessment Target

Whatever their purposes, NSF staff members can design assessments to examine initiatives from three perspectives: (1) the operation and overall effects of the initiative itself, (2) the broader area of investment to which the initiative relates, and (3) the implied models of individual learning or institutional change that the initiative may bring about. These perspectives differ in terms of breadth or depth of focus on NSF's investments.

The three perspectives result from applying a kind of analytical "zoom lens" to the assessment target, schematically shown in Figure V-1. At each level of magnification, assessment can address different topics. By zooming out to view an entire area of investment, NSF's initiatives are seen in the context of other NSF (and non-NSF) initiatives and the conditions that motivate or justify them. From this vantage

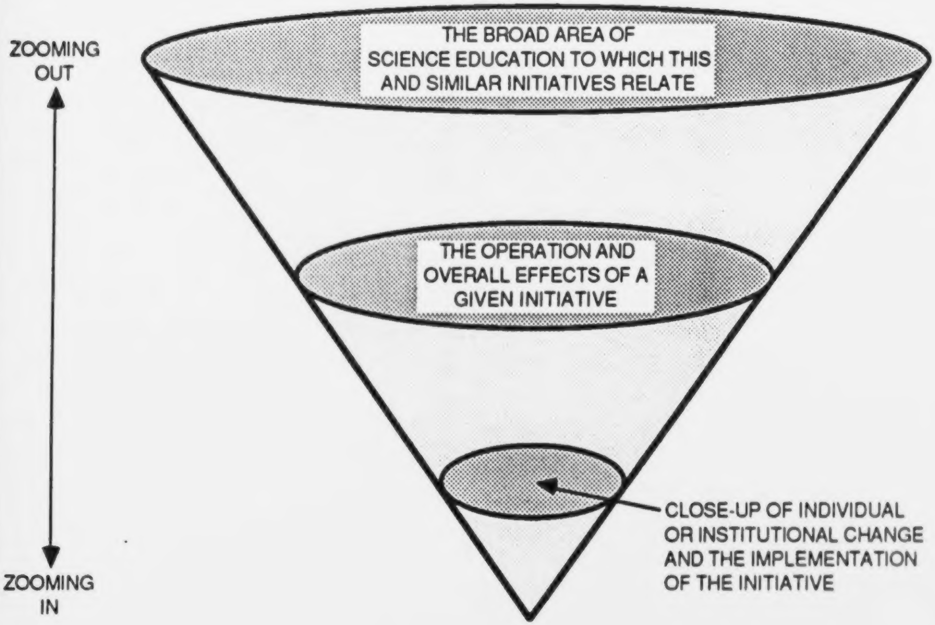
ZOOMING OUT
THE BROAD AREA OF SCIENCE EDUCATION TO WHICH THIS AND SIMILAR INITIATIVES RELATE
THE OPERATION AND OVERALL EFFECTS OF A GIVEN INITIATIVE
CLOSE-UP OF INDIVIDUAL OR INSTITUTIONAL CHANGE AND THE IMPLEMENTATION OF THE INITIATIVE
ZOOMING IN

FIGURE V-1 AN ANALYTICAL "ZOOM LENS": THREE PERSPECTIVES ON NSF'S SCIENCE EDUCATION INITIATIVES

point, the needs of the population as a whole, to which these initiatives are addressed, can be clearly viewed, albeit globally. By zooming in to the process by which individuals learn science (or teachers change their instructional practice), observers get a close-up of the process by which NSF-sponsored activities affect (or can affect) individuals in the science education system.*

The three perspectives provide a framework of topics for assessment, which generate possible questions to pursue. We summarize these assessment topics, stated generically, in Table V-1.

We present our framework in terms of generic topics rather than specific questions for a reason. To take a simple example--NSF's solicitations for materials development through publisher partnerships--there are many productive assessment questions that could be asked about the way NSF has implemented this initiative or how the profes- sional community has responded to it. For example, how do publishers interpret the way the Foundation framed the purposes and requirements of the solicitation? What can be learned from the way NSF has spread the word among relevant segments of the professional community? Do particular combinations of expertise seem likely or unlikely to show up in project teams proposed in response to this solicitation? Many other questions can be imagined, depending on what NSF most wants to know about the proposal solicitation process in this or similar instances, but the generic assessment concern is the same: to examine what NSF does to implement an initiative and the nature of the proposal response to it. To simplify our discussion and to keep attention on the bigger issues in framing assessment questions, we therefore stick to generic topics, rather than try to list all the questions that might be asked regarding each topic.

The Operation and Overall Effects of Science Education Initiatives

Concerns about NSF's investments often focus on the operation of the initiative itself--in particular, on the way NSF carries it out, and how activities funded under the initiative are implemented--and on its overall effects as represented by some aggregate measures of resulting performance, attitudes, or choices of science learners (or other participants). The principal categories of concern--and hence the foci for assessment--are shown schematically in Figure V-2.

The first set of concerns concentrates on **what NSF does to carry out the initiative** (A in the figure) **and on the initial response to these efforts** (B in the figure). NSF may want to get a better understanding of (1) the solicitation process and the way it is interpreted by prospective proposers, (2) the nature of outreach efforts

---

* NSF may also wish a close-up view of other, earlier stages in the chain of events leading to these results--such as the way potential proposers react to program announcements or the ways particular types of projects are carried out.

Table V-1

GENERIC ASSESSMENT TOPICS

Regarding the Operations and Overall Effects of the Initiative

- NSF's implementation of the initiative.
- Professional community's initial response (e.g., proposals).
- Implementation of projects funded under the initiative.
- Types and numbers of participants affected by the initiative.
- Aggregate effects of project activities on participants' or intended beneficiaries' performance, knowledge, career choices, etc.
- Influence of NSF-supported activities on other members of the professional community (i.e., those not funded under this initiative).
- Costs of the initiative in relation to (a) the investment of others (fiscal leverage), and (b) the initiative's overall effects (cost-effectiveness).
- Alternative forms of initiative to address particular needs.

Regarding the Area of Science Education to Which the Initiative Relates

- The relationships among learners, resources, and institutions in this area of science education.
- The nature of the learner or participant population potentially affected by this (or related) initiative(s).
- The presence or likelihood that other resources, initiatives, etc., will be directed at this area (by NSF or others).
- The justification for a federal role and, more specifically, the rationale for NSF's involvement.
- The state of the "infrastructure" (e.g., institutional capacities, professional attention, etc.) in relation to this and related initiatives.

Regarding the Model of Individual or Institutional Change Implied by the Initiative

- How individual learners (or other participants) interact with the learning resources or activities.
- The process(es) of individual learning or change presumed by this initiative.
- The range of individual learning outcomes, including knowledge, skills, attitudes, and choices.
- Variation in the way different types of individuals interact with learning resources and are affected by them.
- (Parallel topics for institutional changes.)

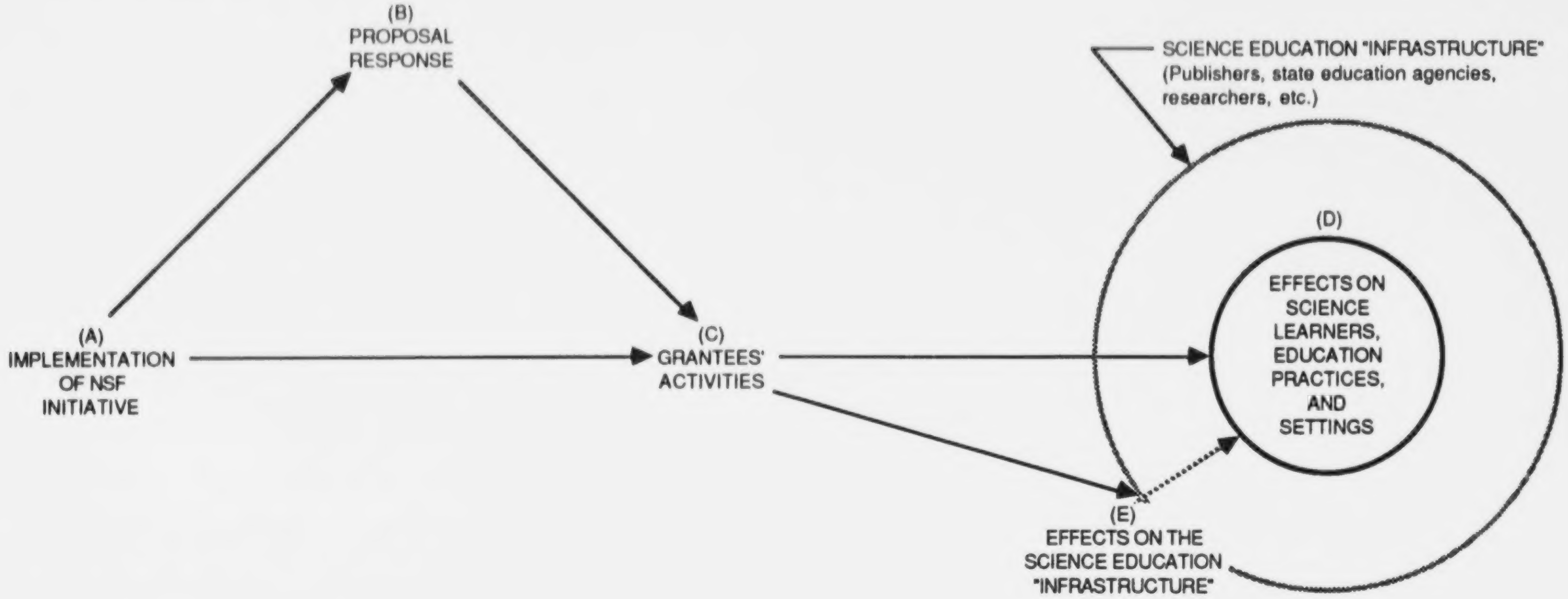
(B)
PROPOSAL
RESPONSE
SCIENCE EDUCATION "INFRASTRUCTURE"
(Publishers, state education agencies,
researchers, etc.)
(D)
EFFECTS ON
SCIENCE
LEARNERS,
EDUCATION
PRACTICES,
AND
SETTINGS
(A)
IMPLEMENTATION
OF NSF
INITIATIVE
(C)
GRANTEES'
ACTIVITIES
(E)
EFFECTS ON THE
SCIENCE EDUCATION
"INFRASTRUCTURE"

FIGURE V-2 MODEL OF THE OPERATION AND OVERALL EFFECTS OF AN NSF-SUPPORTED INITIATIVE IN SCIENCE EDUCATION

and the response to them, (3) the characteristics of proposal teams and the factors influencing their decision to submit proposals, (4) the range and quality of the proposals that are submitted, and (5) the level of resources NSF puts to this initiative. The focus here is on NSF's actions--things that are under its control directly--and the immediate reaction to them by relevant segments of the professional community.

For example, a new kind of initiative like NSF's Private Sector Partnerships to Improve K-12 Science and Mathematics Education (NSF, 1987a) raises many interesting questions about the solicitation process and the response to it. NSF has relatively little experience with private-sector groups (except publishers) as significant partners in improvement efforts; these groups are relatively new entrants into the business of improving science education. How has the highly flexible and open-ended solicitation issued by NSF been received by potential players in the private sector? Are certain kinds of partnership more likely to result than others? What kinds of outreach by Foundation staff seem to generate the greatest interest and most interesting proposals? These kinds of questions could all be asked productively through various forms of assessment.

A second set of concerns has to do with **grantees' activities--in other words, what principal investigators and their teams do once they have received the funding** (C in the figure). Here, NSF may want to gain insight into (1) the range and focus of the activities (what conception of science, instructional levels, etc.), (2) their scientific content, (3) the relationship between these activities and their institutional settings (universities, schools, museums, etc.), (4) the nature of participants in these activities, and (5) the special value or influence of NSF resources in the implementation process (e.g., in attracting matching funds).

NSF's investments in the development of new and comprehensive approaches to middle school science teacher preparation (NSF, 1986b) present an important example for investigating these matters. How do these new approaches address the problem of recruitment? What kinds of teacher candidates participate in the programs as a result? What fusions of scientific and pedagogical content developed for middle school purposes appear to have more general application to other levels of teacher education?

A third focus from this perspective is **the aggregate outcomes in educational settings of the activities funded by NSF** (D in the figure, as influenced by B). In particular, NSF is likely to want information on (1) aggregate effects on individuals (indications of change in learners' or participants' knowledge, performance, attitudes, or choices) and (2) institutional effects (changes in goals, practices, and the use of resources once NSF-funded activities cease). Assessment questions can be framed accordingly. These questions are especially important regarding investments in materials, training of teachers or other "front-line" professionals, and informal science learning resources.

Not all of NSF's science education initiatives aim directly at the science classroom, the teacher, or the learner, however. Some--investments in network formation, dissemination, and knowledge-building, to mention a few--are directed at improving the "infrastructure" for K-12 science education, with presumed indirect payoff for learners in educational settings. For such cases, NSF will want to know about **aggregate outcomes in the professional community** (E in the figure, as influenced by B) and will be particularly interested in (1) the accumulation of new knowledge about science education, (2) the spread or replication of NSF-supported ideas or activities, (3) institutional changes (in organizations other than schools or informal learning settings), and (4) the interaction among professional-community members. For some investments, such as materials development through publisher partnerships, NSF may wish to know about both the effects on educational settings (e.g., how many and what kinds of schools are using a newly developed series?) and on the professional community (e.g., what other developers have taken note of the new series and borrowed from it or imitated it?).

Zooming Out: The Area of Investment to Which the Initiative Relates

The kinds of assessment topics and questions just described ignore the larger context that motivates a given initiative (or any initiatives conceived to address the same needs). By refocusing the assessment lens, NSF can ask questions about the area of investment to which the initiative relates and draw implications for NSF actions, professional response, or the initiative's likely effects. The answers to questions about the area of investment help to establish that NSF's current initiatives are important, do not duplicate what others are doing, and are timely.*

Here, a first set of assessment concerns have to do with **the population of learners (or participants) potentially affected by current NSF initiatives or others that might be designed**. The boundaries for the "population" of interest depend on how the area of investment is defined. We prefer broad definitions such as "informal science learning of children and youth," "school-based mathematics curriculum and instruction," "K-12 science teacher education," and so on, but more precisely defined areas of investment (e.g., elementary science teacher education) could be used as well.

Given a population of interest, such as mathematics learners in school or elementary science teachers, the following kinds of assessment concerns are likely to be important to NSF: (1) size and composition of the population, (2) critical learning needs within the population, (3) its geographic distribution, and (4) how it can be reached.

---

* Much of the analysis we undertook in Phase I was an attempt to answer questions of this sort. There, we selected areas of opportunity for improving K-12 science education, and identified initiatives that were appropriate to NSF by examining investment assumptions at this level (see Knapp et al., 1987b).

A second set of assessment topics have to do with **the institutional infrastructure for serving the needs of the learner population.** Of particular importance are: (1) the current state of knowledge about the learner population, (2) the presence or likelihood of other initiatives and resources directed at the learners' needs, (3) the degree of attention currently given to these needs, (4) the institutional barriers and facilitators to serving these needs, and (5) professional events or trends.

For example, in connection with current or future investments aimed at promoting informal science learning opportunities among young people, NSF may well ask how young people interact with the different channels or media of informal science education. In aggregate terms, which types of learning media (broadcast, museums, print, etc.) capture the greatest portion of time or make the deepest impressions? (We have, in fact, addressed a similar question as part of our Phase 2 pilot test described in Section VIII of this report.) What trends in the development of these media suggest opportunities (or the lack thereof) for NSF to exert leverage?

A final set of assessment topics concern the **justification for a federal role and, more specifically, for involvement by NSF** (as opposed to any other federal agency). Here, assessment questions can address the following topics: (1) the rationale for a federal role, (2) the fit between learner needs and NSF's unique capabilities, and (3) the feasibility of NSF involvement, given its political and resource constraints. Foundation managers may ask, for example: how do NSF's investments in inservice teacher education contrast with and complement (or compete with) those of the U.S. Department of Education? Answers would help the Foundation develop a firmer ground for its own unique contribution to improved continuing education for science and mathematics teachers.

Zooming In: Close-ups of Individual or Institutional Change and the Initiative's Operation

The previous set of topics and illustrative questions seek information about the big picture into which NSF's initiatives fit. But a third perspective on initiatives must also be considered. Do (or will) the Foundation's initiatives make a difference in individuals' lives? In specific terms, how will individual schools, museums, or other institutions be changed by the activities NSF supports? How will the right professional groups be enticed to submit proposals? By refocusing the zoom lens once again, NSF can direct its attention to these questions, examining assumptions about individual learning or change and the way institutions (schools, museums, universities) are affected by the initiative.

From this perspective, assessment questions are framed that provide a "close-up" view. **For some initiatives, questions about the implied model of individual learning and change will be very important.** For example, the design of initiatives rests on assumptions about such things as the way a high school student is influenced by NSF-sponsored science enrichment experiences, the way an NSF-trained "leadership teacher" works with his or her colleagues on returning to school, or the way repeated viewing of Foundation-supported science television series changes a young child's view of science. **In other instances, questions about the fine detail of assumed institutional changes are important--** for example, how a new model for preparing science teachers would be adopted at institutions that did not develop this approach. In addition to asking about the way individuals or institutions are affected, **NSF may also wish to examine more closely how the initiative operates**--for example, by studying the incentives for school-level people to contribute to NSF proposals or by gathering data on individual interpretations of the Foundation's solicitations.

We illustrate the assessment topics appropriate to this perspective by considering the models of individual learning implied by NSF's investments in informal science education. Figure V-3 presents a picture of the informal science learning process presumed to occur as a result of NSF funding. The links in this chain of events suggest the following topics for assessment:

- How the individual interacts with NSF-supported activities or learning resources; the process of learning in the kinds of settings targeted by the initiative.
- The range and type of immediate individual learning outcomes, including the learner's knowledge, skills, behavior, attitudes, and choices.
- The long-range, often indirect effects of these outcomes on the individual's subsequent behavior, attitudes, and choices.
- Variation in the way different kinds of individuals interact with learning resources or are affected by them.

Instead of concentrating on the individual learner, the zoom lens might focus on some aspect of the educational setting, such as the way NSF funding enables science museums to put together informal learning resources. Here, NSF might wish to examine the theory or model of institutional change and would therefore address topics parallel to those concerning individual learners.

Similar topics exist when the Foundation's support does not aim directly at the learning of young people or the settings in which they learn science. Other types of individuals (teachers, administrators) may be the immediate target of NSF's investments. For such instances, categories of questions apply. For example, in the case of teachers, NSF needs to explore how Foundation-sponsored activity influences teachers' motivation for further science learning, science knowledge and skills, images of science, etc. But, in addition, NSF must consider how these activities improve teachers' motivation for improving their professional skills, their grasp of the skills themselves, their images of themselves as members of a professional community, and so on.

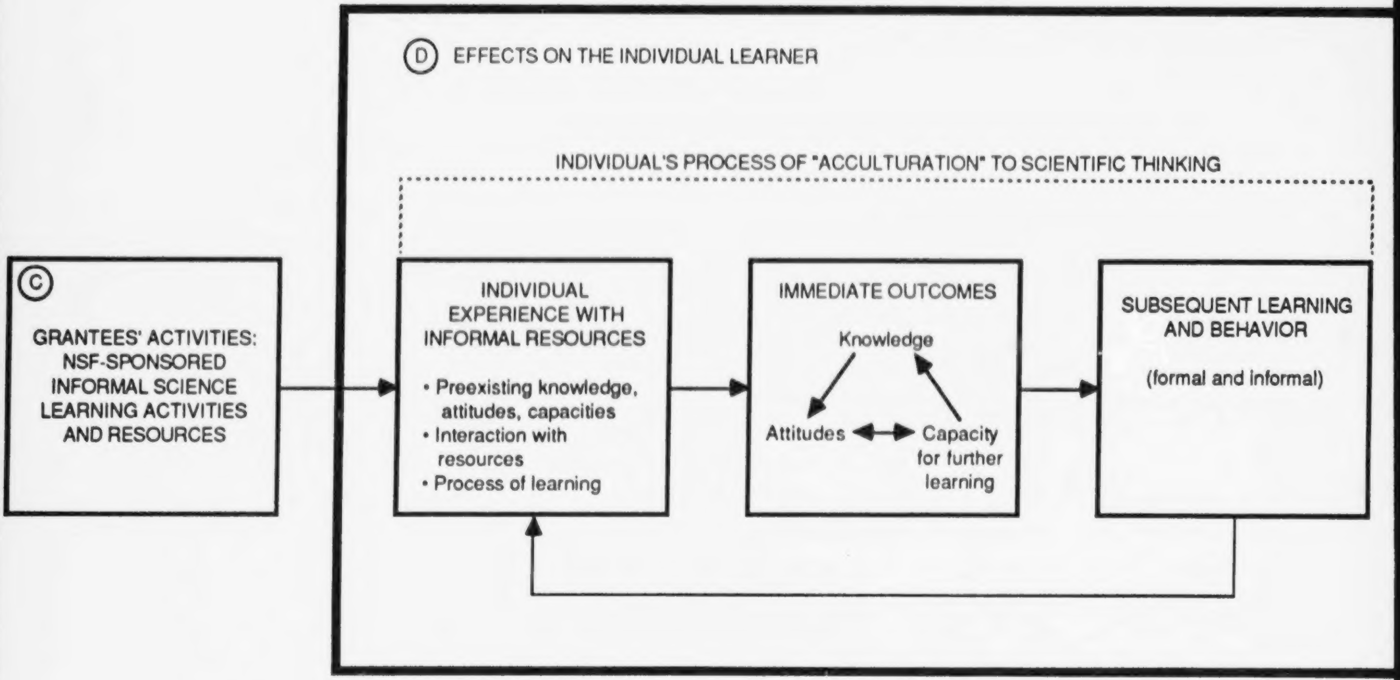
Ⓒ
GRANTEES' ACTIVITIES: NSF-SPONSORED INFORMAL SCIENCE LEARNING ACTIVITIES AND RESOURCES
Ⓓ EFFECTS ON THE INDIVIDUAL LEARNER
INDIVIDUAL'S PROCESS OF "ACCULTURATION" TO SCIENTIFIC THINKING
INDIVIDUAL EXPERIENCE WITH INFORMAL RESOURCES
• Preexisting knowledge, attitudes, capacities
• Interaction with resources
• Process of learning
IMMEDIATE OUTCOMES
Knowledge
Attitudes
Capacity for further learning
SUBSEQUENT LEARNING AND BEHAVIOR
(formal and informal)

FIGURE V-3 MODEL OF INDIVIDUAL LEARNING AND CHANGE IMPLIED BY NSF'S INITIATIVES IN INFORMAL SCIENCE EDUCATION

Whether NSF's investments aim directly or indirectly at individual learners, the most important question from this perspective asks, in effect: is there a plausible (or demonstrated) connection between Foundation-supported activities and changes in the individual learner, teacher, or educational setting? For example, will an intensive focus on science content during a summer workshop prepare "lead teachers" to develop and provide adequate support for their colleagues when they return to school in the fall? Even without relying on a careful documentation of project outcomes, we can answer: probably not. Enough is known about the training process and the transfer of skills to critique the implied (or stated) model underlying this form of leadership training investment. By contrast, a leadership training strategy that emphasizes not only science content but also training in how to cope with school district bureaucracy, diagnose teachers' weaknesses, and elicit school administrators' support represents a more credible approach to the problem. The "theory" behind this latter strategy recognizes forces confronting any attempt to establish a leadership teacher training capacity at the local level. The assumptions still need further examination--for example, by assessing whether training in how to cope with school bureaucracy is transferrable to new situations.

How the Three Perspectives Can Be Used to Document and Examine Initiatives

The three perspectives just described provide complementary vantage points on NSF's science education initiatives. In general, NSF will want to describe each initiative and demonstrate that the logic, rationale, and assumptions underlying each initiative are sound when viewed from all three perspectives. Doing so will help to create the mosaic of evidence about initiatives, which we called for in *Part One*.

We are not suggesting that NSF should gather information about all of the assessment topics just described for each initiative it launches. Rather, depending on the circumstances surrounding each assessment and the "clients'" concerns, certain topics will be important to pursue. Typically, these topics will focus on aspects of the initiative about which NSF knows less or that relate most directly to issues on the Foundation's planning agenda.

It is not a trivial task to arrive at an answerable and important set of assessment questions, but if NSF is to take the knowledge-building function of assessment seriously, it must weigh carefully the assessment questions that really matter the most.

An example illustrates how NSF might consider and select questions to pursue regarding one of its current science education initiatives. NSF's recent solicitation (NSF, 1986a) for the development of elementary mathematics materials that feature the computer and calculator assumes the following chain of events:

> NSF's solicitation and funds will attract leading thinkers and developers in the elementary mathematics curriculum world, who will create prototype conceptions and models of K-6 mathematics education that will in turn inspire or guide curriculum development and teacher education on a wide scale.

The initiative is aimed at a particular need (for better conceptions of the K-6 mathematics curriculum that take full account of the calculator and computer) within a broad domain (mathematics education within schools).

To assess this initiative, NSF would try to examine the assumptions it makes at three levels. Regarding the overall area of investment (K-12 mathematics education in schools), the initiative assumes widespread availability of calculators and computers, inadequate attention to these technologies in mathematics education at all grade levels (or put another way, insufficient attention to them in the early years to build a strong foundation for mathematics in later years), and so on. Regarding the operation of the initiative itself, NSF's solicitation assumes that appropriate proposers are available, established mathematics curricula and teaching approaches are susceptible to change, appropriate groups are able to pick up prototypes and use them, and so on. Regarding the individual learner, the initiative assumes that the new technologies have some intrinsic advantages for students (cognitive, motivational) and are an effective way of learning certain mathematical ideas.

Effective assessment of the initiative would assemble evidence that confirms or refutes these assumptions, concentrating on those assumptions that are problematic, researchable, affordable, and of greatest usefulness for decisions on further NSF investment in mathematical education. There is a reasonably good consensus, based on recent evidence (e.g., California State Department of Education, 1985; Conference Board of the Mathematical Sciences, 1983, 1984; Coxford, 1985; Romberg and Stewart, 1984), that the technologies in question are widespread and that most students in most schools now have, or soon will have, access to them. There is no need to pursue this assumption in any great detail; secondary sources supply sufficient evidence for assessment purposes. (These same kinds of sources reveal other facts about the domain as a whole that complicate the picture--for example, that teachers are generally uncomfortable with these technologies at present.)

There are difficult questions, on the other hand, about the degree to which developed prototypes get noticed and used in subsequent curriculum preparation. This topic would therefore appear to be a better target of assessment resources; however, questions of prototype transfer are extremely complex and difficult to answer. For one thing, with reference to the outcome of current investments, such questions take a long time to answer; important decisions about successive waves of NSF support for curriculum development featuring prototypes will have to be made before the evidence is in. Historical evidence from an earlier era of NSF-funded curriculum development suggests that whole prototypes have not transferred particularly well, whereas pieces of these prototypes have infiltrated the structure of current curricula and textbooks (Quick, 1977; Usiskin, 1985). So what is the question that assessments mounted today can address?

One partial solution would be to assess the process of publicizing and disseminating prototypes under development now, to gauge the likelihood that prototype transfer could take place--that is, how many are printed, who uses them, how do publishers learn about them, etc. Another focus for assessment might be to examine the nature of commercial distribution rights (or the equivalent) for the products that result from current projects, as a way of judging the incentives for prototype transfer. A third focus would be how the ideas emerging from current NSF-supported prototype development efforts are being received by the mathematics education community, as evidence of the wider professional constituency for the products of these investments. Other aspects of the prototype development process might be taken as targets for assessment, but it is important to recognize that, at best, only partial answers will derive from these efforts. The topic verges on the "too difficult" end of the assessment continuum; consequently, NSF would do well to balance its investigation of the prototype transfer question with inquiry into other key assumptions of the initiative, such as whether the grant announcement is attempting proposals of very high quality, whether the size and duration of projects seems appropriate, etc.

The elementary science initiative involving publisher collaboratives, by contrast, is *not* aiming primarily at producing prototypes. Instead, by involving publishers from the outset, the initiative aims to get new and improved science teaching materials into schools relatively more quickly and directly, through established commercial channels. Assessment of this initiative would thus differ in some respects from assessment of the elementary mathematics initiative. In the case of science, questions about the direct impact of the materials, including sales figures, are more pertinent, and one of the key assumptions that is being tested is whether, in fact, substantial change in teaching and learning elementary science can be brought about by involving commercial publishers. We urge the Foundation to document this initiative carefully. Not only is the initiative an especially important, and somewhat controversial, element of NSF's strategy to improve science education, it is also a multi-stage initiative, which lends itself especially well to ongoing assessment. What is learned at the elementary level may be very helpful in future rounds of investment at the middle and secondary levels.

Examining initiatives from multiple perspectives does not necessarily imply that elaborate assessments carried out over long time periods are needed. Thus, one need not wait patiently until all of the currently funded elementary mathematics materials development projects are completed before developing satisfactory answers about the validity of many of the assumptions underlying this effort. Many, if not most, of these assumptions can be examined with evidence from a variety of sources, including (but not restricted to) documentation of the projects themselves. This brings us to the question of procedures for answering assessment questions, which we discuss in the next section.

VI PROCEDURES AND MECHANISMS

To answer the variety of assessment questions described in the preceding section requires a corresponding range of procedures. NSF needs a repertoire of assessment procedures that can handle short-term and long-term informational needs, original data collection and secondary analyses, queries about what is happening in NSF-funded initiatives and about the national needs these initiatives address. To carry out these procedures, the Foundation must establish and make use of appropriate funding mechanisms.*

As we argued in *Part One,* NSF will be best served by carrying out three categories of assessments: comprehensive studies, documentation activities, and short-term focused assessments. In this section, we describe in detail these procedures and mechanisms.

An example presented in the preceding section illustrates how the three types of assessments complement each other. The Foundation's current investments in elementary science materials development through partnership arrangements including a publisher, a developer, and a school system (as trial site) raise important questions about this strategy for improving the science education of the nation's young people. Some of these questions can be answered only by conducting a long-term study of the initiative, for example, to determine whether the involvement of publishers does indeed enhance the widespread distribution of innovative curricula. Other questions--for example, regarding the kinds of matching resources put up by publishers, or the types of trial situations afforded by participating school districts--require more immediate answers because they seek information that can help make mid-course adjustments in successive rounds of funding for this type of project. In such instances, special-focus assessments carried out through site visits or quick phone surveys are more appropriate. Still other questions are best answered by descriptive information gathered by individuals whose task is to document what happens under this initiative. For example, third-party observers working with the project teams might develop descriptions of the collaboration between publishers and developers (a focus of considerable discussion within the science education community), as well as more routine information about the kinds of classrooms, teachers, and students who try out and validate the curricular prototypes.

---

* We purposely keep our discussion at a nontechnical level, although the proposed use of each procedure implies familiarity with technical details (e.g., regarding assessment design, sampling, instrumentation). We assume that NSF's arrangements for carrying out these procedures will include individuals, within or outside the Foundation, who have the relevant expertise.

Because not all of NSF's science education initiatives are as complex as publisher partnerships, the Foundation might choose not to invest its assessment resources in all three kinds of assessment study for each initiative. But across the full range of initiatives supported by NSF, all three types of procedures would be necessary to handle the Foundation's assessment needs.

Comprehensive Assessment Studies

It is easy to conceive of assessment as a large-scale formal "study." That most often means a program evaluation or evaluative study of some kind, sponsored either through a grant or contract mechanism.

Design Options

Comprehensive assessment studies can be designed to gather either prospective or retrospective evidence about initiatives and their effects.*

Prospective Designs--Assessment studies in this mode tend to be designed to assess the achievement of program or initiative goals by gathering information before, while, and after the funded activities take place and subsequently analyzing it to form conclusions about the implementation or impact of these activities. The conventional wisdom among many evaluators is to design evaluation from the start of the program or initiative; studies done "after the fact" are considered weak and undesirable. NSF's new assessment activities conform to this basic pattern, although they differ from one another in some ways. The two recently initiated assessment studies focus on programs (Presidential Young Investigators, College Science Instrumentation) as a whole. Both emphasize early data collection, commencing while (or before) projects are under way. Furthermore, they emphasize all-inclusive rather than selective data collection--e.g., from all of the College Science Instrumentation projects.

There are many variations on this approach to assessment, among them a number of longitudinal designs, but the underlying logic is the same and, in some respects, it is hard to argue with. Studies done in this mode have some obvious advantages:

- *Concurrent Timing.* The studies are especially well suited to capturing information about successive stages in the life cycle of an initiative or program.

---

* Our discussion does not include cross-sectional designs (e.g., large-scale surveys), but we note that these are an attractive option for certain purposes, such as answering questions about areas of investment in science education. Though not conceived primarily as efforts to inform NSF's planning, recent grants for surveys of informal science learning centers (Association of Science and Technology Centers, in progress) and secondary-level science teachers (Weiss, 1988) contribute to that purpose.

- *Comprehensiveness.* These studies appear to examine each initiative thoroughly. Their size and timing permit them to collect data on most aspects of the activities in question. The studies can thus address many (though not all) of the assessment questions that might be asked; typically, they address questions about the operation and overall effects of a particular initiative, but questions regarding broad areas of potential investment or individual learning and change models can also be examined.

- *Credibility.* Because these studies are thorough and comprehensive (and because they are done by third parties), their findings will tend to be given greater weight by external audiences.

- *Visibility.* Large formal studies attract attention and, as such, have the potential to draw large and diverse audiences into the assessment process. For certain purposes that is clearly a virtue, though there are some obvious political dangers.

But for all these advantages, there are significant disadvantages. The first and most obvious is the fact that it typically takes a long time before studies of this kind yield results. For initiatives supporting multiyear projects, assessment findings may not be available for 3 to 4 years from the time that plans for the study are first drawn up. That is a long time to wait for answers. In all likelihood, the results from such studies will not be completed in time either to inform the next set of decisions or to answer the questions of key external audiences (such as federal funding bodies) about this line of investments.

Related to the timing problem is the high cost of conducting such studies. Even by allocating a larger proportion of the Foundation's resources to assessment, as we have argued in *Part One,* NSF will still not be able to mount very many such studies, perhaps one per program per decade at most (assuming an annual outlay of between $200,000 and $400,000 per study). Given the large number of questions NSF is likely to want answered, it is not particularly productive to allocate all the assessment resources to such studies, especially to serve short-term needs. A more balanced allocation of resources to a few large-scale studies and to a larger number of small-scale activities might accomplish NSF's assessment goals more effectively.

In addition, large formal studies with prospective designs are a relatively inflexible vehicle for gathering assessment information. Despite good intentions on the part of those who carry out the assessment and good communication between them and NSF monitors, the designs of these studies--including instrumentation, comparison groups, and data collection and analysis schedules--tend to restrict the collection of information to a particular set of issues and information needs determined at one point in time. The biggest danger is that the assessment will become increasingly unresponsive to NSF's planning agenda as time goes on, with the result that, after years of waiting, the assessment provides answers to questions no one is asking any more.

Retrospective Designs--Retrospective designs present a somewhat underused alternative to the kinds of prospective studies just described. A virtue of such approaches is that they afford the possibility of looking from the "outside in" at the results or consequences of NSF's investments rather than from the "inside out" at the unfolding story of programmatic efforts to reach desired goals. In principle, assessment studies with retrospective designs start with a phenomenon in science education that might be (or has been) influenced by NSF investments and look backward at the various sources of influence on the phenomenon. This form of research assesses investments in reverse order, by starting with long-term outcomes and tracing backward through the chain of events leading to them (Elmore, 1980).

Clearly, for detecting the cumulative effect of influences that are diffuse and long term, although potentially powerful, this kind of approach has its attractions. For examining NSF's investments in informal science learning, such as broadcast and museum exhibit investments, it can provide insight into the residue left by such experiences; in addition, it can shed light on questions about informal science education as a whole, as well as the relative strength and variety of informal influences on individual learning. This approach is also more efficient than prospective designs. By concentrating on the measurable residue of experience rather than the chain of events leading up to, and immediately following, the individual's experience with NSF-supported activities, the study can be done in a shorter time. (By the same token, retrospective designs of this sort are not appropriate for answering some questions, such as ones pertaining to the implementation of NSF-funded projects.) But, most important, this kind of study forces NSF to see the results of its investments in the context of a larger array of influences, of which Foundation-supported activities may be only one.

But the weaknesses and limitations of retrospective designs need to be noted. If undertaken as large-scale studies, assessments with retrospective designs also suffer from some of the limitations noted above for prospective designs, although in lesser degree. In addition, there are well-known weaknesses with retrospective designs (e.g., see Knapp, 1980). Respondents' recall is sometimes vague and inaccurate. The procedure is an inefficient way to learn about the effects of a particular resource (e.g., an NSF-sponsored exhibit). Most significant, the findings from such studies are difficult to interpret. To gain confidence in respondents' own attributions of effect to cause, for example, one must corroborate respondents' accounts with other evidence or probe carefully in exploratory interviews the various influences that might pertain. Typically, this kind of design detects salient influences rather than the fine detail of an individual's learning process over time, but for many assessment purposes that level of detail is sufficient.

Mechanisms for Sponsoring Comprehensive Assessment Studies

The scale and complexity of comprehensive assessment studies imply that NSF must generally secure third-party organizations, through either contracts or grants, to do the work. Contracted studies performed in response to requests for proposals

(RFPs) are the most obvious mechanism; like other agencies, NSF has most often turned to this device when supporting assessments of this type. Although RFP-guided studies have obvious advantages, they also have many drawbacks. NSF should therefore resist the impulse to set up all comprehensive assessment studies through contracts and should actively explore the use of grants as an alternative device.

Contracted Studies--There is an easy rationale for designing assessments (e.g., program evaluations) through RFPs. As outsiders, the contracted parties can provide a more objective account and can bring to bear specialized expertise. At the same time, NSF is able to exert considerable control over the focus and conduct of the assessment activity, especially by the way the RFP is written and by monitoring the contracted work. This kind of control is justified when assessments are designed to answer fairly specific questions about particular types of investments. Finally, by choosing among proposals competing for the same work, the Foundation is more likely (or so the theory goes) to get a good assessment. Indeed, many such assessments *are* of high quality.

But procuring assessment studies through a competitive process also rests on assumptions that may not hold. It assumes that appropriate third parties are available, aware of the procurement, able to undertake the work within NSF's time and cost constraints, and interested in the job. Other major difficulties arise, which parallel the disadvantages of large-scale assessment studies themselves.

- *The timeline for competitive procurement.* Competitive procurements typically take a long time from the initial idea to the delivery of findings or results, particularly if the procured work is set up as a formal study employing a conventional social science research methodology. The cumulative time from inception of the idea to the point at which a contractor begins the assessment work can be close to a year. Add to that 1 or more years necessary to complete most conventional assessment studies, and the total timeline exceeds 2 years at a minimum.

- *Contractual inflexibility.* Although contracts vary in this regard, they tend to spell out in some detail the nature of the work to be performed, the schedule of performance, the methods to be used, and the kinds of products that are expected. The danger is that as time goes on, the RFP's specifications and the project designs set up in response to the RFP become less and less suited to the evolving nature of the assessment.* Especially for longer-term assessment activities, such as studies that span 3 or more years, the risk of becoming unresponsive to important issues on NSF's planning agenda is considerable (although not insurmountable).

---

* Contracts can be modified, as the work proceeds. Our own work on Phase I of this study evolved in significant ways--see the description of study approach and procedures in Knapp et al., 1987c.

- *Staff time and costs.* The many steps in the procurement process, including monitoring the assessment projects, consume a lot of staff time, to say nothing of the costs of the studies themselves (which typically vary from several hundred thousand to more than a million dollars). Understandably, unless the assessment activity is large to begin with, NSF staff may not feel the investment of their time is worth it.

Together, these difficulties make the RFP a cumbersome mechanism for commissioning many kinds of assessment.

We conclude from this discussion that third-party studies supported through competitive RFPs are often not worth the effort. However, the benefits can clearly outweigh the costs, for example, when NSF is fairly certain of what it will want to know several years away or when the size and complexity of the activity requires a third-party study of the kind we have been discussing.

Assessment Grants--Contracts are not the only vehicle for NSF to get what it wants from the outside assessment experts. Grants (e.g., from program funds) can also serve the purpose, although the looser relationship between the Foundation and the third party implied by the grant vehicle changes some of the expectations for the use of this mechanism. Grants are most appropriate for supporting studies of particular initiatives or for encouraging field-initiated work that contributes to the Foundation's overall assessment goals. But because of the length of the peer review process and NSF's inability to direct or specify grant-supported work, this mechanism would be less appropriate for procedures that had a specific short-term assessment goal specified by the Foundation.

On rare occasions, NSF has used grants to support work that assembles evaluative information about its investments. A successful example is a recent study sponsored by SEE's Informal Science Education (ISE) Program of the "3-2-1 Contact!" science television series (Crane, 1987). The proposal for this project went through the normal peer review process along with all other grant proposals to ISE, and was selected on its merits as a reasonable use of program funds. The result has been an insightful and balanced exploratory study of the population this broadcast series reaches and the kinds of short-term effects it has on viewers.

There is no reason why this type of project couldn't be funded more frequently out of existing program budgets. Doing so deviates little from the current grant-making pattern to which NSF staff are accustomed. SEE has already taken a significant step toward supporting assessment grants. A recent program announcement offers an open-ended invitation for proposals to conduct "assessment studies," which are investigations that "address issues related to the ongoing appraisal of the Foundation's many educational programs" (NSF, 1987b). Currently, SEE is placing priority on studies that develop criteria for assessing program effectiveness, identify the characteristics of high-leverage programs, and develop a framework relating national trends to assessment activities; assessment studies on other topics are also welcome. Understandably, this mechanism has yet to generate a substantial response from the professional community; the announcement has not been out long enough to have done so. But it represents a step in the right direction. By combining this announcement with a modest outreach effort to draw attention to this new focus for solicitation, SEE might attract a small number of good proposals addressing key assessment issues.

Other science education programs in NSF have not yet established a pattern of using their own funds for this kind of purpose, although, technically speaking, such investments are permissible under any existing grants announcement. Until the Foundation signals its interest in this kind of work more clearly--in the form of revised grants announcements, individual outreach, or both--members of the professional community are unlikely to think of submitting such proposals to programs that put the priority in grants announcements on topical programmatic goals. This situation may be fortuitous: NSF may well not wish to see a large number of proposals on assessment topics that do not correspond to its most pressing questions. Nonetheless, if NSF staff establish that it is both permissible and important to support such inquiries with program funds *and* indicate areas of assessment interest, their doing so is likely to influence the kinds of proposals NSF receives.

Documentation Activities

A second category of activities generate ongoing descriptive information about what NSF supports. Unlike comprehensive assessment studies, these procedures are designed to assemble a quantitative and qualitative record of the "basic facts" about project grantees, activities, participants, etc. Documentation answers the question "What happened?" in a form that can be quickly and flexibly used for a variety of reporting and program planning needs; these data are also potentially valuable for more comprehensive, long-term assessments. Like assessment studies, documentation activities are best carried out by third-party contractors or grantees, although for some limited purposes in-house staff may do the documentation work.

We review below three types of documentation activities (and associated mechanisms)--documentation grants, ongoing data collection systems, and management information system (MIS) improvements--each of which can contribute a different type of documentation to the Foundation's collective data base.

Ongoing Data Collection Systems

For certain kinds of initiatives, systems can be set up to collect standardized information about grantees, project activities, participants, etc., on an ongoing basis. This kind of system is especially appropriate (1) for initiatives that support the delivery of services to individuals--for example, graduate fellowships, teacher inservice education, science enrichment for able high school students--and (2) as a way of gathering information that is easily counted.

There are various means for creating data collection systems, which vary from simple to complex. In-house data collection systems give NSF more immediate control over data collection and more immediate access to data, but there are severe constraints on the amount and quality of data that in-house staff can gather, given current staff capabilities in this area. Longitudinal tracking of NSF-supported Graduate Fellows or Young Scholars Program participants, for example, becomes extremely difficult to do unless a technical capability (now missing in NSF) is established to carry out this task. More elaborate systems will have to be created and these often require specialized expertise (e.g., in questionnaire design, data base construction). Third-party contractors can be engaged to develop and implement such a system, employing such means as repeated administrations of questionnaires. SEE considered setting up such a system for its Young Scholars Program (NSF, 1987c) and initiated a procurement process for this purpose, but rejected a third-party contract in favor of a more limited data collection effort conducted by Directorate staff.

Documentation Grants and Contracts

Third-party grants or contracts can be issued to support documentation with a more discrete purpose than the ongoing data collection systems just described. Rather than collect standardized information repeatedly, NSF can support small studies that document the activities of particular projects (or sets of projects) in which the process of implementing the project(s) reveals important understanding about a particular problem in science education or its solution.

Other foundations and government agencies (including the Ford Foundation and ED's Fund for the Improvement of Postsecondary Education) have experimented with grants or contracts that support third-party documentation of project activities and results. For example, the 11 projects that are funded by the Ford Foundation to set up collaboratives among inner-city mathematics teachers are being documented by a group unrelated to these projects and funded under a separate contract. In this instance, the Ford Foundation properly recognized that documentation expertise is different from what is necessary to mount a development or training project. Although it is probably inappropriate to do this for all NSF-supported projects, the Foundation could benefit from doing so when (1) a set of thematically related projects are funded at the same time (e.g., elementary mathematics materials development projects) or (2) a particular project is an especially good exemplar of a particular type of activity (e.g., "The Voyage of the Mimi," a ground-breaking example of high-quality multimedia materials development for use in both homes and schools). Here "documentation" includes much of what is thought of under the rubric of "demonstration and dissemination," but the contribution of this activity to answering important assessment questions for both internal and external audiences cannot be ignored.

Improvements in NSF's MIS Capability

The Foundation has begun a long-overdue overhaul of its MIS capabilities and, with some forethought, this revision could be made to facilitate documentation. Of course, the MIS will store and process only information that is routinely gathered as part of proposal review, but even this data is useful for answering a number of descriptive questions about past events, trends, or patterns relating to awards for science education. For example, a program officer may want to know whether NSF has recently made any awards for a particular purpose and, if so, may want specifics (grantee, level of funding, abstract). A division director may wonder how the awards in a particular program break out among colleges, universities, and other types of institutions and whether the pattern has changed over time. The assistant director of a directorate may request analyses of funding by discipline. In each case, an MIS could be of great assistance, making the job of investigating and analyzing NSF's investments less time consuming than searching paper files by hand.*

Recent improvements in the Foundation's MIS include the fact that project abstracts, for the first time, are available on-line--a very important addition. Up to five different funding sources can be listed for a single award, reflecting the fact that different programs often contribute funds to the same award. Also, it is our understanding that a true data base system will be created, cutting across various computer systems with standardized terminology and data elements. We think these are important steps in the right direction. But even with the recent modifications, the NSF MIS is not a particularly flexible system. For example, searching for a particular award requires that the user know the award number *in advance*. Searches cannot be made by such elements as title, topic, or name of the principal investigator. This is a severe limitation.

One possible solution would be to make the data accessible via a more flexible computer program, such as the one we used during Phase I of this study. A conversion might be performed only once a year (for convenience of the Office of Information Systems, or whoever prepares the actual data); even so, the availability of these data for past years in a flexible form would be a great improvement, and converting the data base should not be very expensive or time-consuming. Ideally, searching for key words or phrases in any field (e.g., in the abstract) would be possible.

---

* In our research for Phase I of this project, we performed many such analyses as part of our assessment of what SEE was doing and accomplishing during fiscal years 1984, 1985, and the first half of 1986. Although NSF's MIS had recorded the 500-plus awards made during this time, the system was not flexible enough, and the data in it was not sufficiently extensive, to allow us to use the existing MIS for our analyses. Instead, we created our own data base on an MS-DOS microcomputer using dBASE III Plus (a commonly used data base system). By including such data as the subject matter (discipline) on which the award focuses, the grade level(s) at which it was aimed, and the type of activity (e.g., research, materials development, equipment, teacher preparation), we produced a flexible system that could be searched, sorted, tabulated, totaled, and otherwise analyzed in many different ways. More details on the system we created can be found in our report dated May 20, 1986, entitled "Progress Report: Elaborated Project Plans (Phase I) and Program Funding History."

To illustrate the usefulness of this kind of capability with a current example, a staff working group in SEE is now reviewing awards in mathematics over a period of years. The group would like to be able to search and sort award data in a variety of ways, based on such variables as type of institution, academic department (if pertinent), and characteristics of the principal investigator (such as current title).

The group hopes to answer questions like the following: What is the amount of money provided by SEE for each topical area in science education (e.g., mathematics vs. physics vs. biology), and how does it break down within the field of mathematics (e.g., algebra, geometry)? What proportion of SEE's funding has been provided to schools of education? To schools of arts and sciences (e.g., academic departments, such as mathematics)? What is the ratio of direct to indirect costs for funded projects taken as a whole? Has this ratio changed in recent years, and, if so, how? An improved MIS capability of the sort we have described would make it possible to answer these questions efficiently in the limited time available.

Short-Term Focused Assessment Activities

A variety of procedures complement large-scale studies and documentation activities; these procedures are less costly, quicker, more responsive to ongoing planning issues, less tied to the chronology of funded projects, and more focused on strategic assumptions at both the macro and micro levels. Although not exhaustive, the following categories represent the range of procedures that NSF should consider: (1) limited case studies, (2) quick-response surveys (phone, mail), (3) expert analyses and syntheses of literature or available data, and (4) working seminars (e.g., miniconferences, thematically focused meetings of principal investigators, both within and across programs). We discuss below options within each category; these options are summarized in Table VI-1.

Design Options

NSF has a range of design options under each of the categories of short-term, focused activities.

Limited Case Studies--Full-blown case study examinations of current or recent projects are an expensive and time-consuming form of assessment. A more economical way to derive some of the same insights, sufficient for program planning purposes, is to conduct a limited case study, in which one or a small number of projects are visited for a day, or perhaps longer, depending on the assessment questions and available staff time.

Table VI-1

SHORT-TERM FOCUSED ASSESSMENT PROCEDURES

Limited Case Studies

- Multi-site visits to related projects.*
- Single-site case reviews of critical projects, institutions, etc.

Quick-Response Surveys

- Phone surveys.
- Mail surveys.

Expert Analyses and Syntheses

- "Macro-analyses" (statistical profiles) of an area of investment.*
- Literature syntheses and "white papers."*
- Meta-analyses.
- Market analyses (e.g., of key distribution channels implied by NSF investments).*
- Documentation of key events in the professional community.

Working Seminars

- Thematically focused meetings of principal investigators (for assessment and planning purposes, both within and across programs).*
- Mini-conferences (e.g., to design approaches to difficult assessment questions).*

---

* Asterisks designate procedures included in SRI's Phase II pilot test (described in the next section of this report and in Volume 2).

The basic assessment strategy derives from established traditions in multiple-case research (e.g., Greene and David, 1984; Yin, 1984; Miles and Huberman, 1984). Project sites need to be chosen to reflect the range of local settings addressed by the initiative, not all the settings. By collecting interview and observational data according to a common topical guide, information gathered from each site can be assembled into overall patterns that indicate whether the Foundation's funding assumptions hold up across diverse project settings. This approach parallels NSF's monitoring visits (for example, in some of SEE's programs, staff visit selected projects for a day or two). Limited case studies differ in that they aim at developing evidence related to a particular initiative in a strategically chosen set of projects.

Aside from the fact that they are fast, the obvious advantages of these approaches are that they produce information about the local context for NSF's initiatives and enable some of the subtler underlying assumptions to be examined. Limited case studies are particularly appropriate for answering questions about project implementation, the process of individual and institutional change, or the interaction of learners with NSF-supported resources. At the same time, some questions cannot be answered as well through this kind of technique--for example, questions concerning the long-range impact of research investments on the knowledge base in science education.

Quick-Response Surveys--When breadth of information is more important than depth, quick surveys with relatively small samples are an attractive option. These procedures share the characteristic that they elicit a small amount of information from a number of sites, although the samples are typically too small to ensure statistical generalizability. As noted earlier in this section, surveys may be undertaken as a large-scale formal study, but that is not necessary or even desirable to answer assessment questions such as: In what ways are private foundations attempting to make significant contributions to the opportunities for underrepresented groups in science education? How might these efforts interact with current (or projected) NSF investments in this area? There are not enough private foundations with a large amount of funds and an interest in science education improvement to warrant a large--scale, exhaustive survey. An exploratory phone survey of the 20 to 30 leading private foundations and a handful of knowledgeable observers would lead to a satisfactory answer sufficient for NSF's planning purposes.

Foundation planners and managers should consider two kinds of quick-response surveys: (1) telephone surveys (especially when personal contact is important, open-ended information is desired, and sample sizes are small--e.g., less than 50) and (2) mail surveys (when the above-mentioned conditions do not apply). Both types of survey raise important sampling considerations. We note here only that, more often than not, sampling decisions will need to be made to represent the range of sites, individuals, or institutions relevant to NSF's assessment concerns, rather than to represent statistically a particular population.

Data can be collected efficiently by telephone from a large number of individuals, project sites, or institutions, assuming that the phone interviewers are well versed in the specifics of each site and have a carefully prepared set of questions and probes to pursue. This approach seems more appropriate to examining projects that have already been completed--and in which a conversation with a reflective individual (typically the principal investigator, but others would be appropriate in some instances) would yield lessons learned from that investment approach.

In some respects, this procedure and limited case studies yield similar kinds of information, but the depth and range of information are constrained by the data-gathering approach. Assuming telephone protocols are carefully constructed and the interviewers (NSF staff or others) are reasonably familiar with the initiatives in question, the procedure is particularly effective at eliciting data like project staff reflections on the value of NSF funding, salient features of project implementation, reactions of project participants, and the composition and nature of an area of investment.

The weaknesses of this approach must also be recognized. Respondents are likely to offer information that represents their interests well; although skillful interviewing can probe beneath the surface, the Foundation is always left with one individual's view of the world and interpretation of events. The technique also yields very little local contextual information, except in an interpreted summary form. Finally, the time constraints on phone interviewing limit the number of assessment questions that can be probed effectively through the procedure.

When NSF managers desire information from a larger number of sites in a more standardized form, quick-response mail surveys are more appropriate. Mail surveys are particularly appropriate for data that is countable and easily provided by respondents in a short period of time. In many respects, this procedure and phone surveys elicit similar kinds of information. There are important tradeoffs to be considered, however, and NSF must match its choice of procedure to the particular assessment purpose for which the information is being gathered. The cost of carrying out such a procedure is significantly less than that for phone surveys, but the kinds of information that can be collected are also more restricted.

Mechanisms exist for conducting such mail surveys that NSF might consider. The U.S. Department of Education, for example, maintains a task-order contract for a "Fast Response Survey System," through the Center for Education Statistics. Not only does this provide a useful model (that system is extensively used by policymakers in the Department), it is also a mechanism that may be available to NSF directly, on occasion, through interagency transfer of funds.

Expert Syntheses and Analyses--Instead of examining a few cases intensively or surveying a larger number of cases more superficially, NSF may answer assessment questions by asking appropriate experts to assemble what is known from the available literature and existing data sources. Typically carried out by individuals, these syntheses and analyses are a particularly useful way of addressing questions about broad areas of investment (e.g., what is the size and nature of the candidate pool applying to teacher education programs in science and mathematics? What do the findings imply for initiatives aimed at improving teacher preparation?) and about the model of individual learning and change implied by a particular initiative (e.g., does existing literature suggest how teachers absorb and apply what they gain from one-time continuing education experiences? What models of continuing education appear most likely to influence subsequent practice?). These questions can be answered, of course, by designing studies that collect original data, but such approaches are unnecessary in many instances. Enough research and data exist to answer many assessment questions if this information is effectively aggregated and interpreted by knowledgeable members of the professional community.

Expert analyses and syntheses can take many forms. The following five appear to have particular promise for meeting NSF's assessment needs:

- *Statistical profiles of areas of investment.* By aggregating various sources of available data, analysts can create a portrait of a given population of science learners, the institutions or resources that serve this population, and the kinds of science education capabilities offered by these institutions (see example in Section VIII).

- *Literature syntheses and "white papers."* Because they are familiar with the literature, experts can quickly assemble research and commentary that pertain to a particular assessment issue--for example, the approaches to assessing informal science learning at the individual level (see example in Volume 2, Section VI).

- *Meta-analyses and integrative research reviews.* In areas that have been extensively studied through comparable quantitative research techniques, the findings from a series of studies on a single topic can be synthesized to ascertain larger patterns in the data that answer some kinds of assessment questions (Walberg, 1985).

- *Market analyses.* The techniques and data sources that are commonly used in the private sector for assessing the market viability of commercial products can be adapted to the assessment of current or future initiatives--for example, by appraising the "distribution channels" through which NSF-supported products reach science learners (see example in Section VIII).

- *Documentation of key events in the professional community.* Important gatherings in the professional community often deal with issues that are central to NSF's assessment agenda. Experts participating in these events can brief the Foundation on the relevant outcomes.

We note that NSF staff may be appropriate experts for some such analyses, and in one instance they have a key expert synthesis role to play, which is as yet underutilized. At significant milestones in a program's life cycle or when an initiative ends, program officers or divisional staff can do a retrospective review of the initiative and what it has accomplished. This has happened occasionally in the past, most recently with regard to investments in the public understanding of science (NSF, 1981). These kinds of reviews provide valuable insight into NSF's investments, by taking advantage of the program director's proximity to the initiative, perspective as a program manager, and familiarity with the particular projects funded. This kind of review has some drawbacks, however, which must be considered. Such reviews take time to do well, but if learning from investments is a central goal, as we have argued, then the time is well justified. Furthermore, program officers cannot be expected to be neutral observers of their own activities, but that does not mean they cannot reflect intelligently or critically on the way NSF's original conceptions were (or weren't) realized.

Although there is considerable variety among the procedures we are including within this assessment category, they all share some advantages. First, they are highly economical and quick. Assuming experts are found who already know the relevant literature or are familiar with pertinent data bases, these analyses can be produced by a single individual in a matter of weeks or, at most, several months. Second, they bring specialized expertise (which NSF does not have) to bear on key assessment issues.

This type of procedure faces three major limitations. First, because they are based on existing literature and data bases, expert analyses and syntheses are limited by the quality and extent of these sources. The aggregate data regarding informal science learning, for example, is often incomplete and out of date (see Section VIII); the information about the nature of the museum visitor population rests on evidence collected more than a decade ago for a small number of institutions. Second, individual experts interpret the literature and existing data from perspectives that are based on their disciplinary backgrounds and the directions of their own work. Different experts thus do not always come to the same conclusions about what the literature says. This is not a crippling weakness if NSF turns to experts for insights, not definitive answers; the Foundation can also seek analyses and syntheses from more than one expert on the same topic to maximize the range of interpretations and to identify areas of convergence in expert judgment. Third, for obvious reasons, expert analyses are not appropriate to any assessment question that requires the collection of original data from particular projects or other sources.

Working Seminars--Working meetings of various kinds comprise a fourth category of short-term assessment procedures. These meetings bring together NSF staff, relevant experts, and members of the professional community, who may have been supported by the Foundation, for short (e.g., 1- to 2-day), intensive working sessions to explore questions related to the Foundation's assessment concerns. Such gatherings are especially appropriate when the interaction of different points of view is essential or when contrasts between activities are likely to be informative. Two variations on this theme seem especially appropriate to NSF, the first concentrating on principal investigators' experiences as the primary source of assessment information and the second drawing primarily from expert perspectives on questions of assessment approach.

At best, working sessions of this sort exhibit an important strength: spirited intellectual exchange that can provoke new ideas and insights about NSF's investment strategies and approaches. Such meetings also have a network development function that supports the Foundation's ongoing presence in various investment areas and nurtures interaction among members of the professional community who may not communicate regularly (or at all) with one another. Furthermore, by their nature, working sessions are quick; if well designed and managed, they can cover a great deal of ground efficiently.

But therein lies the principal weakness of such working sessions: they do not have time to work through issues in great detail or to converge on consensus. (This weakness can be remedied by combining the working session with other forms of analytic work, such as individual analyses.) They also are difficult to keep focused on issues central to NSF and its role as a grantmaking agency, because outside experts rarely come to the meeting with the Foundation's perspective. Only by carefully interpreting the variety of views that emerge from such sessions can the most useful implications for NSF be identified.

Mechanisms for Short-Term Focused Assessments

Promising third-party and in-house mechanisms exist for carrying out the short-term activities we have just described. NSF has some experience with this kind of activity, but because small-scale assessment activities have not been used much, the mechanisms are not well established or widely known within the Foundation. We recommend that NSF take steps to enable these kinds of activities, along the lines we describe below.

Third-Party Mechanisms--Although special-focus assessment activities are relatively small and simple, NSF is still likely to rely on third parties for most of the assessment work of this sort. The arrangements for doing so may vary, but all can be designed to share many of the virtues of the competitive RFP: increased objectivity, reduced time demands on Foundation staff (as compared with having these staff doing assessments themselves), and access to a wide range of appropriate technical expertise. Task ordering agreements and personal services contracts appear to be the most useful devices for this type of assessment.

Emulating practices of agencies like the U.S. Department of Education (ED) and drawing on its own experience (e.g., in the Directorate for Scientific, Technological, and International Affairs), NSF can contract with a reputable third party to perform a variety of assessment activities on a task ordering basis. Variously labeled "technical support contracts" and "task ordering agreements," these arrangements put a range of assessment resources at NSF's disposal over an extended period of time, to use on an as-needed basis. Assessment activities are set up as ad hoc tasks, typically funded under the open-ended contract on a fixed-rate or fixed-price basis. A task order can be drawn up, agreed on by NSF and the contractor, and issued in a short time frame, such as a few weeks. Such tasks can cost as little as $5,000 or as much as $150,000 and can be completed on any schedule from a month to a year or more. The arrangement is thus highly flexible. Assuming it has a versatile staff (or access to expert consultants), a contractor supported under a task ordering agreement could carry out any of the short-term procedures discussed above.

Initially, this type of arrangement entails a lengthy procurement process to solicit and secure a good third-party organization to carry out the work. The procurement process resembles that for any RFP procurement described above. Once established, the arrangements we have seen in ED and NSF carry on for a period of years--3 or more before the contract expires. During that time, the process of soliciting and guiding particular tasks is neither time-consuming nor cumbersome.

Several caveats are in order. First, NSF will need to invest time and energy in the beginning to make sure that qualified groups know about the possibility of bidding on this kind of work. Second, the Foundation must use task orders regularly to get the best results from this kind of contract. Contracting firms are happier to enter into these arrangements if their staffs can reasonably expect to be used regularly by the Foundation. For obvious reasons, if task orders are few or sporadic, the contractor's staff will become committed to other work and may not be available when NSF eventually wants them. Third, such arrangements require a substantial amount of monitoring time on the part of NSF staff. This time can be justified, however, by the number and variety of assessment activities that can be completed under the contract. The "monitoring" role can easily evolve into one side of a relationship between NSF and a group of individuals who resemble adjunct staff.

NSF could establish such arrangements in several ways. In the simplest form, a centrally monitored technical support contract might be let, for example, by SEE's Office of Studies and Program Assessment (OSPA) (with or without joint funding by other divisions or offices) to serve assessment requests initiated by any program staff in the Directorate. Alternatively, each division in SEE might establish its own task ordering agreement. The Office of Planning, Budget, and Evaluation (OPBE) in ED operates in this latter mode: it currently maintains four separate "analysis centers" that function as task ordering agreements, each assigned to a different topical territory.

On a much smaller scale, NSF can commission individuals with particular expertise to write papers or perform analyses on topics that pertain to assessment issues. The simplest mechanism for doing so--a "personal services contract" or purchase order totaling up to $10,000--has been used extensively in OSPA as a way of synthesizing what is known about particular topics related to developing a "big picture" of science education: for example, OSPA has recently commissioned papers on such topics as the supply and demand in the precollege science teaching force (e.g., Darling-Hammond, 1987; Oakes, 1987; Welch, 1987). In the past, this mechanism has supported particular analyses contributing to chapters on science education in the Foundation's *Science Indicators*.

This device is simple and inexpensive, and it requires relatively little staff time (assuming that NSF staff already know who the relevant experts are). The mechanism is appropriate for answering certain kinds of assessment questions, particularly those that relate to the "state of the art" and those that require specific, limited analysis of existing data bases.

In-House Mechanisms--In principle, NSF staff can play a greater role in carrying out assessments than they do at present. In addition to designated specialists (e.g., in SEE's OSPA), some program officers have assessment expertise. Many are interested in this kind of activity and have engaged in some forms of assessment already. But at current staffing levels, it is probably not realistic for most NSF staff to conduct assessments themselves. In the Foundation's present mode of operation, grantmaking takes a substantial portion of staff time (more in programs with high "proposal pressure"). In SEE, at least, existing specialists are already heavily committed to a combination of planning work, grantmaking, and assistance to the existing or (projected) assessment projects. Furthermore, it is difficult to imagine most program officers having significant skills in assessment, given the other important requirements for individual capabilities (scientific background, familiarity with schools and school systems, knowledge of the grantmaking process, etc.).

Adjunct staff with skills in science education assessment can compensate for the shortage of NSF staff time or expertise and are well suited to conducting short-term assessment activities. Historically, NSF has taken on various forms of adjunct staff, such as faculty on temporary or part-time assignment to the Foundation as advisors or helpers. Other forms of adjunct staff can be imagined--for example, summer interns or graduate students brought in for specific short-term purposes or fellows associated with NSF on a long-term basis to assist with assessment tasks. Although such individuals are less likely to be useful for overall planning, program management, and grantmaking, they can be especially helpful with particular assessment tasks, assuming they have the right expertise.

NSF might consider such arrangements as the "visiting fellows" supported by ED's Center for Education Statistics (CES): university faculty with particular expertise in statistical analyses are taken on by CES for a quarter or semester to engage in inquiries that are related to the Center's analysis agenda (CES pays the fellows a stipend).* Because prestige and funding are associated with these appointments, high-quality individuals can be brought in for short time periods at limited expense. For assessment activities that require less specialized expertise, such as tabulations of statistical data or quick-response mail surveys, graduate student summer interns might be considered as an alternative.

---

[*] Such arrangements do not differ significantly from the "senior science advisor" role currently in use by SEE.

PART THREE

A PILOT TEST OF SHORT-TERM FOCUSED ASSESSMENT PROCEDURES

Assessment procedures that quickly develop information about a focused question are less familiar to NSF staff than the other categories of procedures described earlier in this report. We therefore undertook a pilot test to demonstrate how a representative set of short-term focused assessments could be used to address the Foundation's concerns.

We confined our pilot test to one area of investment: informal science education. The Foundation's investments in this area raise some of the most interesting and difficult questions for assessment, and NSF was especially interested in focusing on this domain. Because of the diversity and long history of NSF's investments in informal science education, we were able to conduct a range of assessments that drew on a corresponding array of data sources. On the following page, we list the six pilot test procedures that can be classified as short-term focused assessments.* Although they do not exhaust the possible ways to design and conduct such analyses (see Section VI), these activities represent the range of possibilities that NSF should consider.

In this part of the report we describe and interpret our experience with these six procedures, with emphasis on the methodological lessons that might be learned from them. The substantive results of each pilot procedure are reported in *Volume 2: Pilot Assessments of the National Science Foundation's Investments in Informal Science Education*.

We recognize that by confining our pilot test to a domain that differs significantly from investments aimed at formal schooling it may not be so easy to see the applications of these ideas to other areas. In discussing each pilot procedure below, we have tried to suggest how it might be applied to other areas of the Foundation's support for science education.

---

* A seventh pilot activity tested the feasibility of a retrospective study design for identifying scientists' sources of informal science learning. We do not include it in this part because it is not a short-term procedure, but rather a limited pilot for a more extensive study. A complete write-up of this procedure can be found in *Volume 2.*

SHORT-TERM FOCUSED ASSESSMENT PROCEDURES IN THE SRI STUDY PILOT TEST

Limited Case Studies

- A Case Visit Investigation: Assessing Investments in Collaborative Exhibit Development

Expert Analyses and Syntheses

- Describing the Domain of Investment Through Synthesis and Analysis of Secondary Data: A "Macro" View of Informal Science Education
- Market Assessment for a New Investment Area: Examining the Potential for Videocassette Technology as a Vehicle for Informal Science Learning in the Home
- A Literature Synthesis: Assessing the Informal Science Learning Experience*

Working Seminars

- A Cross-Program Principal Investigators' Meeting: Examining Investments that Establish Linkages Between Informal Education Institutions and the Schools
- An Expert Mini-Conference: Exploring the Assessment of Learning in Informal Science Settings

---

* This synthesis was prepared as a discussion paper for the expert mini-conference on the same topic. We therefore do not discuss it as a separate procedure here, although such syntheses are a useful stand-alone assessment product. The discussion paper appears in its entirety in *Volume 2.*

An important caveat must be kept in mind: the pilot activities are not complete studies by themselves. They were designed and executed within a tight time schedule as feasibility tests, intended to illustrate what could be done to address significant assessment questions facing NSF. The findings from the pilot assessments are thus illustrative for the most part. In most instances, assessment activities of this type would need to be carried out with a somewhat greater investment of resources to arrive at more conclusive results (the expert mini-conference is an exception--further investment of resources would not have produced greater convergence of opinion on this difficult topic).

VII LIMITED CASE STUDIES

To test the utility of conducting limited case studies of ongoing initiatives, we carried out an investigation of the NSF-supported Exhibit Research Collaborative (ERC), a joint effort by eight primarily midsize science centers to build and circulate to one another high-quality science exhibits. Although funded as a single project, the ERC resembles multiproject initiatives and thus afforded a manageable way to test an assessment procedure that could be used to study a range of NSF science education initiatives.

Case Studies of Initiatives in Midstream

Assessing the progress of an initiative in midstream raises difficult issues of research design. The ERC example displays these issues in microcosm, and thus provides an excellent case for demonstrating how the issues can be resolved. First, it is a multisite project and each participating institution has its unique culture, capacity, and goals for participating in the collaborative. Second, the project has a number of goals: developing good exhibits, reaching large numbers of people, improving the process of exhibit design, etc. Third, because the project is in progress, any information collected can only approximate the potential outcomes of the project. Assessment results must thus be interpreted accordingly. Finally, the data to be collected cannot be easily or precisely quantified. How did the prototype testing affect the final design of exhibits? Are the exhibits any good? Are there any institutional effects on each center?

These kinds of questions can be most effectively answered through case studies in which multiple methods of data collection are used on-site (interviews, observation, record review). Traditional methods of assessing project progress (e.g., annual project evaluation reports, "show-and-tell" meetings for principal investigators, ad hoc telephone conversations) do not provide the depth of information yielded by full case studies. However, such a strategy is quite costly. In the pilot test, we attempted a modified case study approach, in which costs were reduced by limiting the number of sites visited and our time on-site. Our case study of the ERC can be seen as a test of the cost-effectiveness of limited case studies as an information-gathering mechanism for ongoing initiatives.

Our approach derives from recent work in multisite qualitative evaluation (e.g., Greene and David, 1984; Miles and Huberman, 1984; Yin, 1984). This tradition of evaluation design combines the rigor of standardized cross-site research with the ability to gather subtle, sensitive information about events and processes that are not easily quantified.

Assessing Support for Collaborative Exhibit Development: The ERC Case

Strictly speaking, the ERC is not a formal NSF "initiative" at all, but rather a field-initiated response to the Foundation's program announcement in informal science education. Consequently, our analysis may impute more intentionality on NSF's part than was in fact the case. But for purposes of conducting the pilot test, this assessment target was especially useful for a number of compelling reasons. First, it reflects NSF's philosophy of ensuring high leverage: the project is nationwide, may run for over 5 years, and is intended to reach more than 40 million people. Second, it represents an investment in a category of institution--midrange museums--that has not generally been a target of NSF funds. Finally, the ERC project represents a complex chain of events. By supporting this project, the Foundation is in effect hypothesizing that, through a process of professional collaboration and with the technical assistance of a professional evaluator, a group of disparate institutions will build and share high-quality interactive science exhibits. NSF further presumes that these exhibits, in turn, will provide educationally fruitful experiences for individual visitors in a wide variety of settings.

The project, then, rests on a number of key assumptions; information about its progress--and hence the soundness of these assumptions--could prove useful to both participating institutions and NSF policymakers. Data on the broad national effect of the project (e.g., number of visitors) and its cost-effectiveness would help NSF in reporting inside the Foundation and to interested outside parties such as Congress. Perhaps more important, information on the collaboration among consortium members and their attempts to conduct self-assessments in each of the science centers might help NSF planners to refine future projects or to provide assistance to the ERC members. For the eight participating institutions, information on the progress of the collaborative might assist them to make midcourse corrections or to plan future endeavors.

The Exhibit Research Collaborative

The collaborative has its roots in the demand for high-quality, interactive exhibits in medium-size museums. These museums seldom have the financial and personnel resources to build many high-quality exhibits annually. The ERC was designed as a solution to this problem: with NSF support, each museum focuses its energy and resources on the development of a single exhibit each year but shares in the results of the seven other members.

NSF contributes $1.14 million to the collaborative, while each science center provides an additional $100,000. Using a staggered schedule, each center follows a common design process that includes assessing visitors' knowledge and interests, building and testing prototype exhibits, and creating a finished copy of the exhibit, which will travel to the other museums. Beginning in mid- to late 1987 and continuing through the spring of 1991, each museum receives a new traveling exhibit approximately once every 4 months. Throughout the process, representatives of the various institutions meet regularly to review exhibit topics, assess the collaborative's progress, and hammer out technical details.

The ERC embodies broader goals than the creation of good exhibits that enhance visitors' educational experiences. The project also seeks to introduce staff at these centers to a reflective process of prototype building and testing. This process includes "formative evaluation"--that is, a structured self-assessment to provide information for improving the project as it goes along--which has been at the heart of exhibit development at a number of well-respected science centers (e.g., the San Francisco Exploratorium). A third goal, implicit in the project, is to help the museums build the technical and professional capabilities of their staff.

Procedure for Conducting the Case Visit Investigation

We identified four separate stages in the exhibit design schedule and selected one institution that fit into each:

- Planning stage: Discovery Place, Charlotte, NC
- Prototypes developed: Pacific Science Center, Seattle, WA
- Final exhibit completed: Louisville Museum of Science and History, Louisville, KY
- Received traveling exhibit: Oregon Museum of Science and Industry (OMSI), Portland, OR

In addition, we chose to visit two other participating centers because they were not midsize museums. The Boston Museum of Science is a large, well-established institution, quite different from the other science centers in the collaborative and, arguably, the least in need of participation in such a collaborative. In contrast, the Reuben Fleet Space Theater and Science Center in San Diego is considerably smaller than the other institutions, with greater fiscal, personnel, and physical constraints. Finally, we included a visit to the Science Museum of Virginia in Richmond, Virginia, because its staff evaluator had visited each of the participating institutions and could provide excellent background for our subsequent visits. The sequence of our site visits was driven primarily by geographic considerations and the scheduling constraints of the various museums.

Two staff members took part in the case study, although each site visit involved only one person on-site. Site visits lasted one full day and included semistructured interviews with relevant museum staff, a review of records concerning the exhibit development process, and, where possible, an inspection of prototypes and/or the finished exhibit. Interviewees generally included the director of the museum or other senior staff person responsible for exhibit development, members of the exhibit design staff, fabricators, educators, and, in a few cases, evaluators.

Interviews, document review, and observation of exhibits were structured by a common topical guide. We designed the topical guide to address the following general questions:

- What was the genesis of the collaborative? How did the collaborative form and get NSF funding?
- What is the character of the museum and its community setting? How has this character and setting influenced participation in the ERC?
- Are the project's goals congruent with NSF's goals?
- What is the collaborative actually doing? What kinds of collaboration exist? What kinds might exist?
- How does the exhibit design process work? The formative evaluation component? The collaborative activities?
- What are the staff members' perceptions of the quality of their and others' exhibits?
- What kinds of "outcomes" could the collaborative foster?
- What is the effect of NSF funds on the institution's ability to raise other funds? How does involvement in ERC influence staff capacity and subsequent staff activities?
- How many visitors will actually see the exhibit? Is there any evidence that ERC exhibits have impact on visitors (attitude shifts, etc.)?

Data were analyzed through an iterative process that sought to test tentative hypotheses under a variety of conditions. The goal was not to make summary judgments about each museum's activities, but rather to describe the progress of the consortium as a whole in a way that would help NSF and museum staff refine and improve future activities. Thus, analysis began on-site as we reviewed some of our initial perceptions with museum staff. After each site visit, we wrote short (5- to 10-page) site reports. Those reports specified tentative hypotheses about the entire collaborative (for example: in the formative evaluation process, design staff tend not to adopt the formal, quantitatively based method of evaluation advocated by the technical advisor, but rather adapt their traditional, intuitive evaluative techniques to include more direct input from visitors). We then "tested" the hypotheses at subsequent sites to gauge their applicability under a variety of conditions.

Illustrative Findings About NSF's Support for Collaborative Exhibit Development

Because the ERC is a multiyear project and a number of the centers are just beginning to develop their exhibits, it would be inappropriate to present definitive conclusions about the success or failure of the project. We do, however, offer findings below to illustrate the kind of results that can emerge from a limited case investigation of this sort. A full write-up of these results appears in Volume 2.

The project is on schedule and the collaborative mechanism appears to be working. After some initial scheduling difficulties, the participating centers are following a realistic timetable for the development and circulation of exhibits. Two museums have already finished and shipped their exhibits; two others are finishing the prototyping process; and the other four are in the midst of the development process.

In general, the collaborative is functioning as originally envisioned. The collaborative relationship among participants has developed without seriously hampering the autonomy of each, and the centers are producing exhibits of apparently high quality. To date, the most serious problems have been technical--e.g., involving the durability and adaptability of the exhibits as they travel among institutions.

The collaborative mechanism appears to achieve significant leverage. The collaborative seems to elicit greater effort by museum staff to produce high-quality exhibits; it facilitates the sharing of resources among consortium members; it creates a repertoire of exhibits for medium-size museums that significantly augments their own collections; and in some cases it enhances the fund-raising capability of consortium members. In these ways, this mechanism allows NSF to catalyze exhibit development nationwide in a category of museums it has heretofore not reached extensively.

Many factors affect how each member institution builds exhibits and participates in the consortium. The collaborative functions differently for member institutions, depending on various factors, among them the museum's size (in relation to ERC exhibits), the timing of the project (in relation to the center's own schedule), staff changes, exhibit design philosophy, institutional goals, and political motivations. For example, one science center chose to use the ERC exhibit as the basis for its peak season "blockbuster" and invested an extra $100,000 in the project, while in another museum, the ERC exhibit-building process assumed secondary importance among a number of larger exhibits under development during the same period.

Formative evaluation has had a measurable effect on the design of exhibits. Each center undertook a determined effort to evaluate its exhibit during the design process. Although the style, staffing, and intensity of the evaluation process vary greatly among the museums, all conducted some pretests and all built and tested prototypes before building the final exhibit. In the two museums that have completed exhibits, the formative evaluation effort affected the final design of the exhibits. Such effects included using a different material to stop leaks in a wave tank, eliminating elements of an exhibit that visitors failed to understand, and reconceptualizing the design of an exhibit so that it would appeal to women as well as men.

Science centers face formidable barriers in continuing to use formative evaluation techniques. High-quality formative evaluation requires both a great deal of staff time and a shift of resources from the building of exhibits to the up-front design process. Moreover, it requires specialized staff expertise. Consequently, even the staff who are most committed to institutionalizing formative evaluation are facing a number of practical barriers. One center used Junior League volunteers to help with the prototype-testing process. Another center is trying to raise funds to hire a full-time evaluator. Staff at other centers admit that they will never be able to repeat the formative evaluation of the ERC project without specialized funding for this purpose.

Lessons Learned for Further Application of Limited Case Studies

Our pilot test demonstrated the feasibility of gathering efficient case study information on a complex, multisite project or initiative. In the case of the Exhibit Research Collaborative, on-site visits were the only way to understand fully the progress of the collaborative and the design process in each of the institutions. An accurate description of the formative evaluation process and its effects on the exhibits' design would have been impossible without setting foot in the participating centers. Similarly, interviews with a variety of staff in each center (including, for example, fabricators and educators) allowed us to analyze a wide range of effects on the entire institution.

Just as important, on-site visits allowed us to bring together the perspectives of NSF staff and grantees. Spending a day with an outside visitor helps science center staff rethink the purpose of their work from the perspective of the Foundation. This process of self-evaluation can also stimulate cross-site or cross-project communication. At the same time, site visits help NSF staff understand better the perspectives, needs, and constraints of grantees. This understanding is especially important in the case of initiatives that reflect new or model strategies for meeting goals central to the Foundation's educational mission. Information on the extent to which a model project is meeting these goals and the reasons for its success (or failure) is crucial to NSF program managers as they support projects through multiple years of funding or plan new initiatives. For example, in the case of the ERC, the site visits were able to pinpoint a number of the difficulties museums experienced using formative evaluation techniques for the first time, as well as the barriers museums face in inserting a formative evaluation component in their design process on a more permanent basis. If NSF wishes to support this kind of activity in the future, it may wish to encourage proposers to reshape their formative evaluation plans.

Although it takes more resources to perform limited case studies than less intensive forms of data collection (e.g., phone surveys of the participating project sites), this kind of case study is eminently practical, as the time and cost considerations displayed in Table VII-1 indicate. The cost to NSF of carrying out such a case study (assuming that outsiders do the work) is between 1% and 2% of the total funding for the exhibit collaborative, an amount well within our general estimate of costs for assessing science education initiatives (see Section III). However, because the level of effort required exceeds what NSF program staff currently have available for the assessment function, this kind of assessment activity is practical only if adjunct staff or consultants are brought on to do the job (or if a task-ordering agreement exists to support such assessment activities on an as-needed basis).

The scale of investment in limited case studies can vary considerably, of course. Depending on its purposes, NSF might wish to visit more sites (for initiatives that operate in a large number of projects), spend more time on-site, or produce more detailed write-ups of the results. Any or all of these adjustments would imply greater expense, but the total cost of conducting assessments of initiatives through these means can still be kept to less than 5% of the Foundation's total investment in the initiative. For obvious reasons, NSF is more likely to incur such expenses for larger and more complex initiatives or those that rest on key assumptions that the Foundation wishes to test in anticipation of larger investments in the future. For small-scale or less important initiatives, the need for in-depth information can be met satisfactorily by 1-day monitoring visits made by NSF staff.

In judging whether limited case studies are an appropriate procedure, NSF staff must always weigh the relative value of what case studies produce--detailed qualitative information about a set of investments--against what could be learned from casual contact with principal investigators, monitoring visits by NSF staff, or some form of systematic survey. The benefits of conducting such case studies will not justify the costs when these other means can yield a good approximation of what NSF staff would like to learn. But the Foundation must also consider its purposes in conducting the case studies; to the extent that the information is to be used in formal reporting to outside audiences, a more substantial investment in case study data collection may well be necessary.

Table VII-1

PRACTICAL CONSIDERATIONS IN CONDUCTING LIMITED CASE STUDIES, BASED ON PILOT TEST EXAMPLE

	Case Visit Investigation of the Exhibit Research Collaborative
Number of sites and duration of visits	6 project sites; 1-2 days each site
Time scale, from time of negotiation with NSF till completion of written summary	4 months
Products (see Volume 2)	Cross-site written summary
Resources	
(a) SRI professional staff time	9 person-weeks
(b) NSF staff time	--
(c) Estimated cost*	$18,000

---

* Cost estimates assume that the assessments are conducted by an outside group at a rate of $75,000/professional person-year (plus incidental expenses for travel, secretarial support, etc.). NSF staff time for discussing assessment activities and reviewing results has not been figured into the cost estimate.

VIII EXPERT ANALYSES AND SYNTHESES

We chose for the pilot test three expert analyses that differed in breadth of focus, types of data source, and relationship to current initiatives.

The first, a "macro" picture of the informal science education domain as a whole, presents a statistical profile of the domain as a way of describing the context for NSF's investments, especially in science museums and children's television. This analysis was based entirely on available data from a variety of sources.

The second, a pilot market analysis of the potential for NSF to invest in videocassette technology for informal science learning at home, examines consumer demand and the nature and strength of the commercial channels by which NSF-developed products in this area might reach a mass audience. This analysis was done as a pilot market survey, using techniques developed for the private sector.

The third, a synthesis of literature pertinent to assessing informal science learning, focused on how the individual's experience with informal science learning resources might be conceptualized and studied. Here, we drew on various traditions of research and assessment to build a model of this subtle and elusive learning process and to suggest ways for studying or evaluating these effects.

We describe in this section the first two analyses. Because the third was a preparatory step for one of the working seminars described in Section IX, we do not discuss it at length here, although we offer some general observations about it and other types of expert analysis in our concluding remarks.

Describing the Domain of Investment Through Synthesis and Analysis of Secondary Data: A "Macro View" of Informal Science Education

The purpose of this activity was to develop a statistical portrait of informal science education, both from the perspective of members of the public (especially students up to the age of 18) who are engaged in a great many activities providing informal education and from the point of view of the institutions that offer and support it. Primary attention was given to the roles of television and museums, because these institutions are especially important in science education generally and in NSF's current funding strategy for informal science education. Had time permitted, we would have expanded our research to include more detailed information concerning print media, as well as zoos and aquaria. Also, we would have had a variety of experts review the available data to help us sort out reliable from unreliable or inadequate information.

This "macro view" sketches the big picture of activities in the United States involving informal science education. It is intended to help managers at NSF answer such questions as:

- Outside of work and school, how do people spend their time?
- Which informal activities and institutions are especially important to people (especially young people)?
- What is the role played by informal institutions in contributing to public education in the sciences?
- How do NSF's investment priorities for informal science education correspond to the resources expended by the public and by other agencies?

Potentially, the task of developing a "macro view" can contribute in many ways to planning, reporting on, and justifying NSF's activities involving informal science education. Data synthesized in the macro view help to illuminate not only the domain in general, but also the roles of some specific "agents" and their impacts on individuals. Thus, the macro view can be useful for viewing initiatives from any one of the three perspectives described in Section V.

In some respects, the macro view is similar to the National Science Board's *Science and Engineering Indicators*, but on a far smaller scale. Like that publication, the macro view synthesizes data from many sources both to illuminate specific aspects of a large-scale social enterprise and to describe the context in which it operates. The purpose of both efforts is "to inform national policymakers who must allocate resources to these activities" (National Science Board, 1988).

This macro view has been constructed using existing data sources. Original research is generally unnecessary for this purpose and would be far more expensive. We did not look for data that emphasized NSF's specific role for two reasons. First, the intent is to paint a picture of a domain hundreds of times larger than NSF's contribution (at least, as measured in dollars). In addition, the audience for this product--Foundation managers--is generally familiar with the role of the Foundation and with specific documents (e.g., budget justifications) that document NSF activities and their impact.

In carrying out this task, SRI and NSF are not simply focusing on informal science education but are also testing the feasibility of developing a type of statistical portrait (the macro view) of the domain to which one or more NSF programs relate. By zooming out from specific NSF activities to focus on a much bigger picture, statistical portraits such as this contribute information that is not often part of the manager's day-to-day work, yet is important in grounding specific NSF activities in a larger context. This particular effort seems in many respects a typical example. In developing macro views for other domains pertinent to NSF, for example, analysts would encounter similar constraints, such as the fuzzy definition of "informal science education" or limited availability of data. If the pilot task demonstrates that a macro view is feasible and useful in this instance, then the domain for other Foundation programs or initiatives might profitably be the subject of similar research in the future. Two candidates of special interest would be macro views of teacher education in the sciences and of K-12 science instructional materials. Each of these domains is central to one or more NSF programs, and each is sufficiently complex and diverse to warrant a statistical portrait of the sort we sketch below.

How We Constructed the Macro View of Informal Science Education

To be of maximum use to NSF managers, a macro view of a particular domain should be brief and readable, focus on descriptive information (as contrasted with opinion or speculation), and include a large amount of hard data (with sources noted). In the case of informal science education, such informal institutions as television and museums are so fundamentally different that a decision was made to break the overall task into pieces corresponding to different media.

The basic approach used was to proceed from the bottom up--that is, to collect a great many factual items and data tables first, and then use these to construct a picture of the informal science education field. In practice, there was often an iterative process at work, in which the existence of certain data would help fill in the picture, but would also underscore the existence of a "hole" that needed to be filled with data not yet gathered. For example, knowing the ratings of a number of science shows on public television led us to wonder about the ratings of commercial science shows, and we then proceeded to fill that particular "hole." Whenever possible, a general-purpose statistical reference such as the *Statistical Abstract of the United States* (U.S. Department of Commerce, 1986) or *The Condition of Education* (e.g., U.S. Department of Education, 1987) makes an excellent starting point for research, precisely because such sources often focus on "the big picture."

A great many different sources of data were tapped, beginning with the shelves and file cabinets of the researchers.

- A variety of libraries were used--for example, in the case of museums, the Smithsonian Museum Reference Center, reference material at the headquarters of the Association of Science-Technology Centers (ASTC), the GWU Gelman Library, and several of the Stanford University libraries. A majority of the libraries used are now catalogued on computer (or compact disc, in one case), making searches faster and easier than in the past.

- Discussions with experts in the field proved useful, both for preliminary research and to answer specific questions. Staff at the Smithsonian Museum Reference Center, for example, were able to provide access to numerous documents in response to our requests for a general orientation and for visitor surveys. The directors of research for the Corporation for Public Broadcasting and for the Public Broadcasting System were very helpful in answering our requests for certain specific data relating to public-television viewing.

- Numerous documents were obtained especially for this research from sources in several different states. For example, results of a Field Institute/ California poll were provided at low cost after the nature of our research was explained. The Public Opinion Laboratory at Northern Illinois University provided us with a variety of useful reprints.

Illustrative Findings

Our write-up of the macro view identified various features of the informal education domain, which we grouped under three topical headings: the public and its use of time, television and informal science education, and American museums as a source of informal science education. (Because of the limited time available for this analysis, we did not pursue in much detail other important media or channels of informal science education, such as print or recreational activities; a more complete macro view would have included such topics.)

Regarding the public and its use of its time, we found that:

- Excluding time for work (or school) and sleep, Americans, on average, put at least a third of their total weekly time into activities (television, reading, crafts) that can involve informal science education.
- Approximately one-fifth of a national sample identify leisure-time pursuits with a significant scientific component as their most important informal learning activity.
- Orders of magnitude can be assigned to the amounts of time Americans attend to different informal science media: for example, an estimated 60 times more hours, on average, are devoted to public TV science viewing than to visiting a science museum.

Second, regarding television as a source of informal science education, we found that:

- American audiences watch 20 hours of commercial television, on average, for every 1 hour of public television.
- Approximately three-fifths of the public that are "attentive" to science policy, two-fifths of the "interested" public, and relatively few of the "noninterested" public regularly watch science shows on television.

- A very high proportion of children's programs on commercial television involve at least one theme or aspect explicitly and unambiguously related to science (including space or science fiction). Nonscience television programs, such as dramatic series and the news, convey a great deal of the public's information and misinformation about science.

- Programs about animals, science, and nature are a highly valued part of the public television schedule.

Third, regarding the availability and use of science museums, we found that:

- Science museums are increasing in number and are extremely popular; they are visited by numbers far out of proportion to their representation in the museum population.

- About half of science museum visitors are children. Data about the composition of the visitor population are extremely weak (and NSF is supporting survey work that will partially remedy this situation).

The basis for these and other characteristics of the informal science education domain appear in the full write-up of the macro view, which appears in Volume 2.

Lessons Learned for Further Applications of the Macro View

Despite the preliminary, broad-brush nature of this particular macro view, we were able to find many data sources that provided pertinent, useful information. This pilot exercise was sufficient to demonstrate that the production of macro views can be useful to NSF for the following purposes:

- *As an orientation tool.* The macro view provides the Foundation with a potentially useful orientation tool for many people: new managers; senior personnel whose responsibilities include many different domains; managers interested in the given domain, but whose principal responsibilities lie elsewhere (e.g., in the case of informal science education this might include managers of research programs); and--perhaps to a lesser extent--current managers of the program most closely involved with the particular domain (who may already be very knowledgeable about the domain). Peer reviewers, in some cases, might appreciate having a macro view available.

- *As input to the design of funding strategies.* A number of questions are raised by the macro view that may provoke NSF staff to consider variations in the design of funding strategies. For example, given the importance of newspapers and other print media as a source of general information (presumably including information about science), should there be a special, ongoing role for SEE's Informal Science Education Program in this area?

- *As a way of demonstrating that funding assumptions are sound.* A picture of an investment domain such as informal science education can help to verify that assumptions underlying a given initiative are sound. To take an obvious example, data from our macro view on the nature of the young viewing audience and the time it spends in front of a television set confirm that NSF's current funding for science television aimed at young children is appropriately targeted. The large proportion of science content conveyed by commercial television might prompt a renewed search by NSF for ways to support development that can be picked up by commercial channels.

- *As an indicator of new areas for research or studies.* "Holes" in the data about a particular domain may signal important areas for new research or studies, possibly supported entirely or in part by NSF.

The macro view we produced has significant limitations, however, due in part to the quality of existing data and in part to the constraints on our time and resources. A more rigorous and complete macro view of informal science education would have included more attempts to cross-check and interpret "suspect" statistics, as well as external review of the analysis by expert consultants. For example, relying on industry figures alone for estimates of public-television viewership is weak, because publicly available figures from these sources serve many functions, including promoting public-sector television. A more thorough analysis would have enabled us to contrast estimates from different sources and adjust accordingly. To do a more rigorous review, of course, requires more resources, but because a broad domain of science education may encompass many areas of NSF's investment, the effort to understand the domain may well justify the expense.

Because statistical profiles of this sort are so dependent on the availability of usable aggregate data, the different types of "holes" that can be found in the data deserve further comment. We see three types of holes. The first and most obvious are those data that are clearly needed for painting a picture of the domain, but that are of very poor quality or are missing entirely. An example would be many types of data about museums, such as numbers and demographics of visitors; as our review points out, these data are terribly out of date (see Volume 2). (A current NSF-supported survey by the Association of Science-Technology Centers will soon produce much valuable new data that can help to fill this hole.) Also, there seems to be increasing recognition within the field of museum evaluation that data about what museum visitors learn is inadequate.

A second type of "hole" appears when an aspect of the domain is not treated at all in the macro view, simply because of the institutional slant (or set of questions) used when performing the research and synthesis. Our macro view says nothing about how much time adults, especially parents, spend with young people on various science-related activities out of school. A variety of data can be found to help illuminate this aspect of the domain, such as data on the small amount of "quality time" most parents spend with their children--e.g., fathers spend only an average of 8 minutes a day on weekdays reading, conversing, or playing with their children (ISR

Newsletter, 1985-86). It is not necessarily a simple matter to know what questions to ask about a domain, so that some "holes" will probably always be present.

The absence of explanatory information creates a third type of "hole." Much of the data describing a domain is simply descriptive: who, what, where, etc. But for NSF, it is very important to understand, to the extent possible, why things are the way they are in a given domain. Answering this question often requires looking at studies and types of data that are not conventionally included in a statistical profile of an investment domain. In the case of informal science education, for example, one might look for information that illuminates why the public's understanding of science is as poor as it is. The following kinds of questions might be considered: Are some types of scientific and technological information very threatening to some groups of people and are therefore resisted by the media or the public? How are widely held naive theories or misconceptions about science inhibiting better public understanding of science? Why and under what circumstances do people tend to mistrust or ignore expert scientific advice?

Intelligent synthesis of existing research and information from available data bases can suggest some answers to these questions. Such knowledge comes from a combination of sources, including survey research, psychology, social psychology, sociology, and other disciplines. In preparing a macro view, special efforts may be needed to gather and synthesize data bearing on the question of why the domain is as it is.

A Market Analysis of a New Investment Area: Examining the Potential of Videocassette Technology as a Vehicle for Home Science Learning

This pilot analysis tested the utility of carrying out preliminary market assessments in areas of potential NSF investment. In this case, we looked at the market for home-based use of science-related videocassettes, focusing primarily on teenagers and adults as users. (The definition of "science" is more difficult for young children, which is one reason we excluded them from the analysis.)

There were several reasons for performing this task. First, we were responding to a specific inquiry from SEE staff: to bring information to bear on the question of whether NSF should consider supporting major projects and/or an initiative focusing on home-based videocassettes. The converging trends of decreases in time devoted to science-related TV viewing and increases in the use of VCRs led to speculation that an NSF initiative of this sort might be appropriate. Many ideas for new initiatives surface within the Foundation, but each must be judged in light of relevant information and experience to determine both feasibility and a degree of priority. Expert analysis and synthesis of data (including market assessments) can contribute to this process.

The second reason for conducting the pilot analysis was more general: to better understand a rapidly evolving aspect of the informal science education domain. This domain is relatively complex, involving many institutions and media, and it changes relatively quickly. We identified the use of videocassette recorders in the home as one key area of change in recent years. The technology has rapidly assumed an important place in American culture (note that it is already present in more than half of all American households), yet there is much that we do not know about its potential as a vehicle for informal science education--a fact that led us to seek additional information.

The conduct of either a preliminary or full-scale market assessment seems, in general, useful to NSF for either or both of these purposes--that is, either to further develop the general understanding of an investment domain or to test ideas for specific initiatives against disciplined inquiry. We are using this particular pilot task to explore the utility of both of these applications of market assessments.

Description of Procedures for Conducting the Market Assessment

SRI International has performed market assessments for many clients, and we used established market research procedures to conduct this one. The objective of these assessments is not only to size the market, but to understand its characteristics: the underlying dynamics that forecast growth, stability, and so on. The primary tools are secondary research (using a variety of information and data sources); interviews of key individuals knowledgeable about aspects of the market in question, or closely related markets; and, if warranted, the conduct of a broader survey of consumers or producers to test preliminary findings. A market assessment necessarily combines facts and data with judgments about such information.

For NSF, we conducted a preliminary market assessment based on these procedures. Results of our preliminary assessment are presented in the same overall format as would be used by SRI for a full-scale market assessment. The latter, however, would be considered more reliable because much more data would be gathered.

Initial leads and information came from several sources, including trade publications, key individuals in computer software publishing companies, and their counterparts in allied educational media firms.

- *Billboard*, a trade magazine, publishes weekly information about best-selling videocassettes; and several articles provided important references, for example, to specific educational materials and to individuals in the industry.

- More than one computer software publisher has examined the potential market for prerecorded videocassettes, and thus knowledgeable individuals in this industry were able to provide us with names of other key individuals to interview.

- Early in this effort, the *Video Sourcebook* was identified as an important source of information about developers and distributors. Together with cross-checking obtained via the interviews, this is the source through which we identified most of the firms listed in the more detailed write-up appearing in Volume 2.

Lengthy interviews were conducted with about a half-dozen individuals, and shorter ones with several others. Those interviewed included the head of the education division of a major distributor of educational videocassettes; three people at the vice-presidential level whose firms market various instructional media (especially computer software and filmstrips); and several individuals involved in the textbook publishing industry, including a former vice president who is now a consultant to the industry.

Interviews focused on a number of pertinent topics. These included some description of the processes by which products are selected for development, developed, and marketed; identification of very successful examples of science-related videocassettes; ratings of companies involved, or potentially involved, in this market; and others.

Throughout the pilot task, our focus was not simply on "the bottom line" (e.g., rating the potential size of the market), but on understanding the dynamics of the industry and the environment in which this market is and will be developing. This is approach is a standard feature of market assessments, and directly parallels our advice to NSF that, in general, assessments should produce a greater *understanding* of the topics in question, not simply quantitative information.

Illustrative Findings

A full description of this pilot activity can be found in Volume 2. Here, we simply touch on several major findings:

- *Currently, the videocassette market to provide informal learning in the home is at the embryonic stage.* Development of a significant market niche for these materials appears to be approximately 5 years away, or more.

- *Normal market forces seem unlikely to produce a significant increase in market size for at least 5 years.* The preliminary assessment did not uncover any special barriers to market development that might be reduced or removed by NSF (and thus providing a special rationale for the Foundation's involvement). This would not preclude NSF from supporting exploratory research and development to demonstrate the most effective forms of VCR-based learning activities for the home.

- *Schools are making a steadily increasing use of videotapes for instructional purposes. The use of VCRs in school is likely to have a spin-off effect on the home market for instructional videotape over time.* (In some respects, this situation is parallel to the developing market for instructional computer software.) NSF may want to examine the school market for instructional videotapes in more detail.

Lessons Learned for Future Application of Market Assessments

Market assessments do seem to be a useful tool for investigating potential Foundation initiatives. In the case at hand, a reasonable conclusion is this: to the extent that NSF is interested in widespread market penetration, an initiative in this area would not be appropriate at this time. A related area (the school market) may be worth investigating further. To the extent that it wishes to demonstrate the further potential of VCR technology for home science learning, a modest level of exploratory research and development could be justified. It is precisely these types of judgments that NSF needs to make in considering any potential new initiative.

The conduct of a preliminary market assessment also seems a useful means for obtaining more information about the informal science education domain. Specifically, an understanding of the use of VCRs in the home is important, because this equipment has become ubiquitous and is accounting for a significant amount of time in typical households. Similar market assessments in other science education domains might also provide useful information to NSF.

This was a preliminary market assessment. As such, it should be supplemented by other pertinent information, since it is less reliable than an assessment based on far more data. Fortunately, in this case, the preliminary assessment seems to confirm opinions based on other evidence.

After completing the pilot study, we wondered whether it would be useful to modify slightly our usual procedures for conducting market assessments to focus on the role of government (or other nonprofit) agencies. This could be done through the interviews with key individuals, with the expectation that they might provide important information about the role of these agencies (if any) in overcoming market barriers. If this focus were added, we, and NSF, would need to be sensitive to potential bias from respondents who might have strong feelings, either pro or con, about involvement by government agencies in commercial marketplaces.

Reflections on the Further Use of Expert Analyses

The preceding examples are only two of many types of expert analysis or synthesis NSF may wish to support in assessing its science education initiatives. Other types, briefly noted in Section VI, differ in the kinds of data that form the basis for analysis and the reliance on formalized analysis procedures (such as meta-analysis).

Whatever the type of analysis, each can be carried out efficiently and at low cost. Typically, NSF need only find a single analyst; few logistical arrangements are necessary, by contrast with other categories of short-term procedures such as in the convening of meetings or the collection of data through limited case studies or surveys. The costs and time scales for expert analyses, as demonstrated by Table VIII-1, make them practical and feasible for answering assessments on a quick-turnaround basis. We note, however, that, as with any focused assessment, NSF may invest more or less, depending on its purposes. A full market assessment, for example, would have cost 2 or 3 times what we spent to explore the videocassette field. Similarly, a more rigorous and complete macro view of informal science education would have taken more resources than what we indicate in the table. But, in relation to the scale of investments that might be influenced by these assessments, the costs can be fully justified.

As a class of assessment activities, expert analyses are thus both flexible and efficient, because they rely on information that is already gathered and often internalized by the expert analyst. But the reliance on existing knowledge is also the principal weakness of this type of assessment: it is limited by what is already known or readily available in a form that can be analyzed. For example, our synthesis of literature related to the assessment of informal science learning was limited by the paucity of work in this area. (NSF may still wish to consult expert opinion in such instances, as we did in the expert mini-conference described in Section IX, but the perspectives offered by participants must be recognized for what they are--opinions rather than analysis.) Such analyses are also restricted to areas in which appropriate experts exist--in particular, individuals who have knowledge of the area in question, good analytic skills, and a good feel for the perspective of a federal grantmaking foundation.

One further weakness needs to be considered. Expert analyses are typically carried out by a single individual and, as such, are likely to reflect the biases, preconceptions, or disciplinary background of that person. No matter how qualified or respected the analyst, NSF may wish to verify the outcome of a single analysis task in one of several ways: by commissioning different experts to conduct parallel analyses on the same or similar topics, by conducting informal peer reviews of analysis findings, or by coupling the analysis with another activity, such as the working seminars described in Section IX (we did just that in preparing a synthesis of literature regarding the assessment of informal science learning as a discussion paper for a mini-conference on the same topic).

Table VIII-1

PRACTICAL CONSIDERATIONS IN SUPPORTING EXPERT ANALYSES, BASED ON PILOT TEST EXAMPLES

	A Macro View of Informal Science Education	Pilot Market Assessment of the Potential for Videocassettes in Home Science Learning
Scope of data sources reviewed	▪ Available national data bases ▪ Literature on science television viewership, etc. ▪ Literature on science museums, audience, etc.	▪ Interviews with company executives, industry observers ▪ Industry literature
Time scale, from negotiation with NSF staff through completion of written summary	▪ 3 months	▪ 2 months (3.5 months[*])
Products (see Volume 2)	▪ Written statistical profiles of (a) informal science education; (b) television and informal science education; (c) museums as a source of informal science education	▪ Pilot assessment write-up ▪ Listing of firms ▪ Design for more complete market assessment
Resources		
(a) SRI staff time	▪ 7 person-weeks	▪ 5 person-weeks (15 person-weeks[*])
(b) NSF staff time	--	--
(c) Approximate cost[**]	▪ $14,000	▪ $10,500 ($30,000[*])

---

[*] Estimate for a full-scale market assessment; SRI's pilot was only a feasibility test for such an assessment.

[**] Cost estimates assume that the assessments are conducted by an outside group at a rate of $75,000/professional person-year (plus incidental expenses for travel, secretarial support, etc.). NSF staff time for discussing assessment activities and reviewing results has not been figured into the cost estimate.

IX WORKING SEMINARS

To test the feasibility and usefulness of working seminars as an assessment device, we conducted two meetings that differed in the assessment questions addressed, the experts who participated, and the relationship to issues of assessment design.

The first, a meeting of principal investigators funded under different NSF science education programs, examined the development of linkages between schools and informal education institutions. Participants included the directors of projects located in science museums and other institutions such as zoos or arboretums, along with representatives of five NSF science education programs, each of which supports (or could support) projects that establish such linkages. In the meeting, the project directors pooled their experiences in creating connections between their institutions and the schools; in addition, they discussed possibilities for future NSF investment in this area.

The second, a mini-conference of individuals (including several NSF program officers) expert in the assessment of informal learning, explored issues related to assessing what is "learned" by individuals who interact with informal science resources such as museums exhibits or television shows. By contrast with the first meeting, the mini-conference was aimed at determining how assessments of individual informal learning should be done, rather than producing assessment "findings."

Together, the two meetings illustrate the key role that members of the professional community can play in the assessment process, both at the design stage and later, as information from project work is informally synthesized to gain insight into important planning matters. Meetings such as these add a reflective component to NSF's support for science education, by bringing a variety of expert perspectives and project experiences to bear on questions related to the Foundation's funding strategy. NSF might undertake a variety of such meetings in answering other questions about its support for science education.

A Cross-Program Principal Investigators' Meeting: Examining Support for Projects That Establish Linkages Between Schools and Informal Educational Institutions

The purpose of this pilot activity was to explore ways that gatherings of NSF-funded principal investigators can be used to answer assessment and planning questions. Furthermore, we designed the activity to examine an area of investment that does not correspond clearly to any one of the existing NSF programs. That way, we hoped to encourage NSF staff to take on a more strategic view of investments across grant program boundaries. In addition, we wished to demonstrate that informal, impressionistic assessment of initiatives in midstream could contribute to thinking about areas of investment that were not currently a designated priority for Foundation funding.

Support for projects that create linkages between formal and informal educational institutions was an ideal topic for the pilot. Such investments represent an area of considerable promise as a target of new initiatives (see Knapp et al., 1987b). Although it has not been an explicit goal of NSF funding to date, a number of projects funded over the last 5 or more years have created some form of linkage between informal educational institutions and the schools--for example, through teacher training based in informal education institutions, or materials developed by these institutions for the schools. There was thus a good deal of experience on which to draw.

Assembling groups of principal investigators is a familiar procedure in some NSF science education programs. For example, in the last 4 years, principal investigators managing projects in teacher enhancement, teacher preparation, and studies of science education have gathered for small, regionally based meetings to report on the progress of their respective projects and to share information that would be useful for further work in each project. So far, such gatherings have not been used to answer questions or develop information about issues on the Foundation's planning agenda, but there is no reason why this cannot be done.

This procedure yields evidence that is impressionistic and anecdotal, but if the participating principal investigators are systematically chosen, assessment topics are explored thoroughly in the meeting itself, and the results carefully interpreted (e.g., in a written synthesis of the meeting's proceedings), this kind of evidence can contribute considerably to SEE's understanding of its current and potential investments.

NSF has convened few meetings, if any, to examine promising areas of investment that straddle program boundaries. In such cases, assessment activities that encourage interaction among program staff and project directors who are contributing to a common area of investment can be particularly helpful. At the least, participants can become aware that their disparate projects share a common goal and approach for improving science education. Better still, they can consider whether NSF should make the implied strategy behind these efforts a more explicit initiative.

Procedure for the Meeting

We undertook such a "meeting of minds" by gathering NSF staff and principal investigators of projects that create some linkage between their own informal science institutions and schools. Our goal in selecting participants was to represent (1) the range of projects funded to date that have contributed to this investment area and (2) all SEE programs currently or potentially supporting such projects. We also tried to include diverse settings in which such projects might occur; we recognized, however, that most existing projects of this type are found within larger urban areas, where major informal education institutions are situated. To maximize the number of linkage arrangements represented in the meeting, we decided not to include representatives of the schools in addition to the principal investigators from informal educational institutions. Other criteria figured into the choice of participants as well: we included individuals with recognized standing in the informal science education field, who were articulate and thoughtful, and whose perspectives were likely to differ from one another.

In total, eight individuals from informal science learning institutions and five NSF staff attended the 1-day meeting in which we searched for lessons from project experiences that might inform future efforts to carry out this kind of investment. We did not compensate participants for their time (however, we did reimburse them for travel expenses). Apparently, the topic itself, the chance to interact with colleagues, and the opportunity to help shape NSF's thinking about support for science education were sufficient motivators. We documented the meeting discussion; the write-up of results in Volume 2 of this report interprets the implications of the day's activities for future investments.

Illustrative Findings

The meeting generated a range of ideas about the possibilities for fostering linkages between informal science education institutions and the schools. We grouped these ideas under five categories: (1) the range of existing linkages, (2) the types of barriers to linkage that must be overcome, (3) promising entry points for establishing stronger relationships between formal and informal education institutions, (4) caveats regarding the formation of linkages, and (5) advice regarding NSF strategy. We review below highlights of the findings to illustrate the kinds of information that can arise from such a meeting; a complete write-up of meeting results appears in Volume 2.

Range of Existing Linkages--The eight project sites exhibit a diverse array of connections between informal education institutions and the schools, far richer than one might suppose from knowing the NSF-funded project's goals. These linkages take a number of forms, in particular:

- Organized use of science museum resources by groups of children.
- School personnel and students assuming working roles such as institutional "associate" positions within the science museum.
- Teachers receiving training or support of various kinds at the informal educational institution.
- Institutional personnel working with teachers and classes on school premises.
- Materials developed by the institution for use by the school.

- The establishment of museum-like learning centers or resource rooms within the schools.

- Formal institutional connections at the budgetary and policymaking level.

Typically, the institutions represented in the meeting had established a number of these linkages simultaneously.

Types of Barriers That Must Be Overcome--Meeting participants identified several critical barriers to linkage that must be overcome if a durable and productive relationship is to exist between schools and informal education institutions. Perhaps most important, the "two cultures" need to be bridged--that is, school people need to appreciate and value informal science learning as a legitimate mode of education, and at the same time, informal institution people need to appraise more accurately the goals and constraints inherent in the formal educational system.

Curricular and instructional policies, often formalized in state testing and requirements, pose a second and related barrier. School people often have difficulty visualizing how informal science learning modes can help them meet these requirements; however, in some states, recent increases in requirements (e.g., for science instruction at the elementary school level) have brought educators to the door of the informal educational institution looking for help.

Other significant barriers explored during the meeting included the unwillingness or inability of schools to commit resources (such as release time for teachers to attend training events) that would support a relationship with informal institutions, logistical problems (e.g., transportation to and from a science museum facility), and the limitations on the physical capacity of informal institutions.

Promising Entry Points--Entry points discussed at the meeting derived in part from the unique configuration of events, people, and opportunities in each institutional setting. However, depending on the informal institution's chosen role vis-a-vis the schools--for example, as a repository of unique intellectual and physical resources, a safe haven for professional renewal, or an agent of change in the school curriculum--three entry points seem especially promising in a variety of settings:

- As a neutral arena in which science and education are intertwined, informal institutions can establish long-term supportive relationships with individual teachers (and, to a lesser extent, students)--for example, by employing these people in "museum associate" roles or through other means of professional development and renewal.

- Informal institutions are in an excellent position to play an intermediary role between universities and the schools, by bringing together the resources (both scientific and pedagogical) of the former and helping to translate these into terms that are useful to practicing educators.

- Informal institutions are especially well suited to the development of creative curricula that expand the school's repertoire for experiential science learning.

Caveats in the Formation of Linkages--Although meeting participants were generally enthusiastic about the importance and possibility of forming linkages with the schools, they pointed out grounds for proceeding with caution. Significant trade-offs exist when fostering these relationships. For example, the more closely museum exhibits or activities are tailored to existing curriculum, the greater the risk of compromising the essential spirit of informal learning and discovery. In exploring linkages with the schools, informal education institutions need to consider carefully where the "center of gravity" of their efforts lies--closer to the schools and their current curriculum or closer to the informal institution and its own program structure. In so doing, the informal institution must not compromise its unique strengths.

Another kind of caution concerns the type of clientele informal education institutions can and do reach in their efforts to establish linkages with the schools. Meeting participants recognized that disadvantaged populations, often located in the inner city--and the school systems serving them--are generally harder to bring into long-term and meaningful relationships with the informal educational institutions, although it is easy enough to attract individual students to museum exhibits and activities. These segments of the community would therefore require extra attention, effort, and, possibly, specialized strategies to engage in linkages.

Advice to the Foundation--By interpreting the remarks of participants, we were able to suggest implications regarding the Foundation's degree of focus on linkages, the adequacy of its current program structure for supporting work in this area, and the possibility of NSF's assuming a greater advocacy role in promoting the concept of linkage between formal and informal educational institutions.

- *The sharpness of focus on this area of investment.* Rather than targeting specific types of entry points (e.g., teacher associate roles, traveling kit design), NSF is better off establishing a broad and strongly stated goal of fostering linkages between the informal institutions and the schools.

- *The adequacy of the current program structure.* Most promising activities for establishing or improving linkages between informal science education institutions and the schools can be supported under existing NSF programs. Given this fact, it is probably unwise to consider radical alterations in existing programs. However, unless the Foundation sends clearer signals to the field about its interest in this area of investment, relatively few proposals are likely to arrive that take the establishment of linkages as a central goal. NSF can signal its interest by such means as aggressive outreach to potential proposers, altered priority statements in program announcements, or adjustments to the review process.

- *The possibility of an advocacy role for the Foundation.* NSF has the option to adopt a more visible posture in promoting linkages between informal education institutions and the schools. Apart from what it does to attract and fund proposals in this area, the Foundation can try to project one (or more) vision(s) of the relationship between schools and informal institutions--through position statements, commissioned papers, networking, and conferences--as a way of orienting members of the professional community toward possible actions in this area.

Lessons Learned About Principal Investigators' Meetings as an Assessment Tool

Although not representative of all the ways to focus principal investigators' meetings on assessment purposes, this activity underscored several lessons about the use of this procedure.

Natural incentives for participation make it easy to convene such meetings but hard to stimulate a critical examination of assessment issues. A major motivation for people from the field to participate in such a meeting was undoubtedly the chance to interact with representatives of different NSF programs. For NSF staff, motivations varied, but probably included the desire to get perspective on an area of investment related to their programs and, perhaps, to get a break in the routine of processing grant proposals. These motivations can make it more difficult to achieve the meeting's purposes: first, individuals from the field may try to use the meeting to "sell" themselves to NSF staff, and second, the NSF staff may attempt to solicit proposals conforming to their current definitions of programs. Selling oneself and soliciting proposals are both legitimate functions, but they have little to do with assessment and planning. These motivations tend to make the exchange of ideas uncritical, unless steps are taken to facilitate a more penetrating assessment of issues.

The format of such meetings is exceedingly flexible, permitting discussion to range freely but also creating a problem of focus. This feature is particularly useful for addressing questions about investment areas that are relatively undefined, as was the case in this meeting. The flip side of flexibility, however, is a lack of focus; it is hard for individuals who do not interact regularly to coordinate their thinking enough to generate focused responses to NSF's assessment concerns. All too easily, such meetings can disintegrate into a series of individual agendas competing for "air time." The major challenge, then, is to allow the participants' differences to be expressed yet at the same time to frame the discussion so that issues are joined in a productive way. This is especially difficult when the meeting is restricted to a single day; one participant left our meeting wishing out loud that the event would continue because it had only just reached the point that solutions to the more difficult issues were beginning to emerge.

In selecting points of view to be included in the meeting, NSF must confront difficult trade-offs, particularly if it wishes to keep the working seminar small. For this meeting, we chose to invite the individuals most directly connected to NSF, who were almost all education directors (or staff) in their respective institutions; as such, they tended to lack the perspective of the informal institution as a whole (executive directors would have brought that), and, as we noted earlier, they represented only one side of the linkage relationship. Although the resulting discussion was productive, it did not deal with questions regarding linkages at the institutional level, NSF funding strategies, or the responsiveness of the schools and their possible roles in partnership with informal institutions.

The results of such meetings must be carefully interpreted to yield clear guidance for the Foundation. Because discussion does not typically reflect the federal grantmaker's perspective, remarks must be interpreted in terms of the Foundation's mission, capacities, and funding strategies (as we have tried to do in our write-up of the meeting's results--see Volume 2). For this reason, we strongly encourage that these meetings be designed with some means of generating a formal synthesis of results--either by a third-party documentor or an NSF staff person (conceivably, a nonparticipant principal investigator could play this role, but NSF would have to look hard to find an individual with the requisite breadth of perspective).

The kind of meeting we convened, and most other forms of principal investigator meeting one can imagine, maximize breadth of coverage over depth. Because of the number of participants, the time they take to learn "where each is coming from," and the differences in their viewpoints, much of the time is spent raising possibilities and responding to each other's ideas. Accordingly, no one project's experiences are fully or systematically examined in this kind of setting; rather, they are selectively tapped to provide illustrations, rationale, or counterpoint to the ideas that are under consideration. The meeting is thus a good way of "brainstorming" possibilities, but at the same time a weak method for plumbing the depths of a given project's experiences.

In summary, this kind of procedure is particularly good for extracting lessons from project experience and for developing alternative interpretations of that experience. The interchange between individuals who represent different kinds of investment and who do not normally communicate with one another helps to accomplish this goal. Necessarily, the amount of information gained about any particular project is more circumscribed, and it is virtually impossible to standardize the information across projects, unlike in case studies or surveys.

An Expert Mini-Conference: Approaches to Assessing the Effects of Informal Science Education on Individual Learners

This pilot activity--a mini-conference on assessing informal science learning--was aimed at examining how individuals interact with NSF-funded informal science education resources and what they learn from those interactions.

The issue of individual learning is important to NSF because the justification for its investments presumes that people who use informal education resources gain something educationally valuable from them. Finding appropriate assessment approaches and methods for testing that belief, however, is anything but straightforward. There are several reasons for the difficulty. First, federal support of informal science education resources (museums, television, etc.) aims at a wide range of loosely articulated educational and cultural goals. Second, there is no well-established theory of informal science learning on which to base assessment questions and approaches. Third, informal science learning experiences are very different from one another and from formal learning experiences. For all of these reasons, the practitioners of informal science education (as well as NSF program officers) are extremely skeptical about using many assessment approaches that derive from the formal learning domain.

For NSF, then, the task of assessing what people learn from the informal education resources that the Foundation funds is both important and difficult. Unlike other pilot activities described in this report, no simple illustrative study would add to NSF's knowledge of how to do this kind of assessment in an ongoing way. Instead, we felt the need to "back up a step," to gain a larger perspective on the issue of assessing informal science learning and to try to find general approaches that would (and would not) be useful for NSF to use. This need for a critical review of past and current assessment approaches and for a deep rethinking of the assessment task made the topic of assessing informal learning a good candidate for a small working conference of experts.

Our working session was designed to provide an opportunity for Foundation staff to explore this particularly difficult and important assessment issue with the best minds in the field. NSF program and division officers, caught up in the daily pressures of processing proposals, rarely have the chance to spend a day or two exploring fundamental questions of Foundation strategy or policy, especially in the area of assessment. Even more rarely do they find the opportunity to involve a range of experts in their deliberations. Thus, our working seminar sought to illustrate a mechanism by which NSF program and division officers could find an arena in which they might reflect on larger, long-term issues.

Designing and Conducting the Expert Mini-Conference

The process of choosing and inviting these individuals required considerable time and effort, including much interaction with NSF program officers, literature review, and networking.

For the meeting we brought together experts deliberately chosen to represent diverse fields and perspectives on assessment. The participants included a physicist with long experience as a science educator and author; an elementary science specialist with extensive experience as a film and book reviewer; a political scientist specializing in the study of scientific literacy; a physicist and specialist in cognitive studies in science museums; an art museum administrator who is also an art historian and museum educator; an evaluator, expert in inquiry-based science learning; an applied educational researcher specializing in children's television; a communications and marketing researcher; and a museum exhibit designer who had conducted a great deal of research on exhibits. Attending the meeting from NSF were program officers from SEE's Informal Science Education and Research on Teaching and Learning programs, and divisional staff from the Division of Materials Development, Research, and Informal Science Education.

In conducting a meeting like this, a delicate balance exists between chaos and order. On the one hand, the meeting must be structured, have well-articulated goals, and be guided to keep it from degenerating into a discussion of issues that may or may not be related to NSF's primary interests. On the other hand, prematurely constraining the form of discussion, outlining the exact nature of the solutions desired, or demanding a consensus where there is none limits the seminar participants' ability to explore issues fully.

To provide structure for the meeting, as well as to introduce a common framework for the discussion, we prepared a "discussion paper," which was distributed to all participants before the meeting. This paper (see Section VI in Volume 2) outlined three questions as the focus for the discussion:

(1) What kinds of learning are most important in informal science education? In posing this question, the paper outlined in a schematic way the logic of NSF's informal science investments as an influence on individual learning.

(2) What assessment approaches and procedures can be brought to bear in assessing these outcomes? The paper reviewed past and existing approaches to studying informal science learning.

(3) On which of the possible assessment procedures should NSF concentrate its efforts? In raising the question of priorities, the paper discussed other factors to be kept in mind--such as the different audiences for assessment information and the differences among informal education media.

At the meeting itself, SRI staff served as facilitators, moderating the discussion and keeping it focused on the issues central to NSF. The meeting lasted 1-1/2 days, with half-day sessions addressing each of the points above. This format worked particularly well because the overnight break provided a chance for informal but important interactions, and allowed SRI staff to summarize the first day's discussion and present the summary to the groups for revision on the following morning. This process of summarizing, feeding back, and clarifying previous discussion allowed the group to participate more fully in the formulation of the meeting's findings.

The results of the meeting were presented in two forms (in addition to this discussion of the procedure): an interpretive summary of conclusions (see Volume 2, Section V) and a reconstructed dialogue of the meeting (see Volume 2 appendix).

Results of the Mini-Conference

Out of the meeting's discussion, the following general guidelines emerged, which can shape future assessments of what individuals learn from informal science education experiences.

Informal science education should not be thought of in the same way as formal science instruction in schools. Despite an extensive literature on the unique nature of the informal learning environment, the tendency of those involved in assessment is still to understand the purpose, activities, and outcomes of informal learning in terms of concepts derived from formal education. Determining appropriate assessment methods or underlying philosophies in the informal domain requires a different framework of concepts that have yet to be developed.

The central mission of informal science education is acculturation to the scientific world, not the teaching of specific content or skills. Becoming scientifically literate means becoming more familiar with, and more a part of, the "culture of science, mathematics, and technology." Thus, informal science education investments can be seen as efforts to contribute to the acculturation to the world of science, mathematics, and technology.

There are several advantages to using the idea of "acculturation in the sciences" as an overarching goal for informal science education. This concept (1) helps those engaged in assessment look beyond short-term knowledge or attitudinal "gains" from informal science experiences; (2) reinforces the idea of learning as the interactive, cumulative experience with science in both formal and informal settings; and (3) connotes a lifelong process of developing interest and knowledge in science as well as becoming comfortable with scientific habits of thought. The notion of acculturation is highly compatible with NSF's overall mission of broadening the pool of people who are competent and interested in science (see Knapp et al., 1987b).

Assessment should explore and document the ways in which informal science education resources contribute to this acculturation process. Five guidelines seem especially important in this regard:

- *Documentation--both statistical and qualitative--should play a large role in all assessment efforts in this area.* Given the complexity of the informal science learning experience, it makes sense to focus assessment first on answering the question: what is happening in informal settings?

- *The value of NSF-supported informal education resources should not be judged solely or primarily on the basis of the empirical evidence that people "learn" from them.* Meeting participants agreed that it is a mistake to assess informal resources as if they were the main source of cognitive learning about a phenomenon or the main determinant of attitudes about science and mathematics. Much of their impact may come through complicated and subtle interactions with many other sources of information. Informal learning resources may thus contribute to acculturation without having a single or main impact.

- *To capture the cumulative and long-term nature of the acculturation process, NSF should experiment with methodologies that measure the long-term impact of experiences in informal settings.* Longitudinal studies of the developing interests and skills of young people may help shed light not only on the role that informal resources play, but also on the interaction of school and out-of-school experiences. Retrospective studies (e.g., see Section VII in Volume 2) may help uncover common patterns in the development of scientific interests and talent, and help understand how scientific interests are either nourished or discouraged at early ages.

- *Key projects should be studied intensively.* To complement the broad (and low resolution) view of retrospective and longitudinal studies, several key projects could be assessed much more closely with an eye toward documenting and understanding the processes of interaction and the impacts of the projects.

- *There may be an important complementary role for expert judgment and criticism.* A collection of criticisms from a range of experts, in combination with a statistical understanding of the numbers in the audience that are "reached," may provide a better understanding of the learning opportunity provided by NSF-funded informal resources than any empirical measure of individual learning outcomes.

Overall, a program of applied research is needed to search for new ways to think about, describe, and approach the assessment of informal learning. What the field requires now is the development and articulation of a broader and clearer rationale for the investment of public money in informal science resources. It is premature to talk of the effectiveness or the cost-effectiveness of investments in this area, since the overall goal of the enterprise is not well understood. The attempt to describe informal learning as an important part of a larger acculturation process is but one example of the kind of rationale building that is needed. Assessment efforts can help in building this rationale, first, by articulating broader visions of the enterprise and, second, by describing the process and outcomes of informal learning.

The need now is for a program of applied research that pursues these areas. Progress in assessing (understanding) informal learning depends as much on work that helps to develop and articulate an overarching view of informal science education--its nature, mission, and role--as on the assessment of particular NSF-funded activities. Better theory, a meta-analysis of the work in the field, and assessment paradigms appropriate to the new formulations of the enterprise are needed.

Lessons Learned for Future Use of Mini-Conferences for Assessment Purposes

This working-group seminar was particularly good at exploring issues of learning and assessment that are difficult to articulate, but it was less successful at producing a consensus about NSF assessment policy or practices. It did little to offer detailed technical solutions to assessment problems. The lack of consensus and the degree to which the meeting focused on assessment policy rather than technical issues appeared to bother some of the participants, who wanted to arrive at more specific and concrete suggestions for NSF. Others, including the NSF representatives, were happy to have the freedom to probe larger, more abstract issues and to generate a more general framework for thinking about assessment in this domain.

More specific lessons can be gleaned from this experience:

- *It appears to be important that a third party organize and conduct the meeting.* Not only does this relieve NSF program officers of the time-consuming job of organization, but it also puts the meeting on neutral territory, where NSF staff and outsiders can participate as individuals equally interested in the problem.

- *The structuring and facilitation of the meeting are crucial to its success.* The discussion paper put everyone on common ground at the beginning, and the models introduced in the paper served as useful springboards for discussion. Also, if the seminar is to focus on a substantive issue, then it is important that the facilitator be very knowledgeable about the topic and able to ask the right questions to further the discussion.

- *Taken together, the three products of this seminar--discussion paper, meeting summary, and reconstructed dialogue--are an effective way of communicating and interpreting the thinking of meeting participants.* Although requiring more effort than the usual "minutes of the meeting," this three-part reporting approach could be a useful model for similar meetings in the future.

Reflections on Further Use of Working Seminars

We have already commented on lessons learned from each working seminar, but several overall observations deserve mention. First, although they differ in complexity and depth, these meetings represent only a small investment of resources, as demonstrated by Table IX-1, and can be organized in a fairly short time frame. They are thus a practical approach to certain kinds of assessment questions. We note, however, that the more elaborate form of working seminar exemplified by our mini-conference requires a significant amount of staff work. Under current conceptions of their role, NSF program officers would be hard put to manage that kind of effort; therefore, third parties or adjunct staff brought in for this purpose would be

Table IX-1

PRACTICAL CONSIDERATIONS IN CONDUCTING WORKING SEMINARS, BASED ON PILOT TEST EXAMPLES

	Cross-Program Principal Investigators' Meeting on Linking Informal Institutions and Schools	Expert Mini-Conference on Assessment of Informal Science Learning
Meeting size and duration:	12 participants (5 from NSF); 1-day meeting	12 participants (3 from NSF); 2-day meeting
Time scale from initial negotiation to write-up of results	1.5 months	3 months
Products: (see Volume 2)	▪ Meeting summary	▪ Discussion paper ▪ Meeting summary ▪ Reconstructed dialogue of the meeting
Resources:		
(a) SRI professional staff time	3.5 person-weeks	10 person-weeks
(b) NSF staff time	7 person-days	5 person-days
(c) Estimated cost*	$6,500	$20,000

---

[*] Assuming the meeting was conducted by an outside group at $75,000/professional person-year (plus incidental expenses for travel, secretarial support, etc.). NSF staff time for attending the meetings and reviewing meeting products has not been figured into the cost estimate.

necessary to make the seminar successful (for most such seminars, a qualified facilitator could be easily secured through a personal services contract).

Second, these meetings do not "speak for themselves." Although the interchange within the meeting has important residual influence over the thinking of participants, the "results" of the meeting must be constructed after the fact and interpreted to make them maximally useful to NSF. By themselves, such seminars produce a collage of ideas; efforts at securing greater consensus during the meeting are not only doomed to failure in most instances, but probably counterproductive. A separate effort must be made by an NSF program officer or appropriate third party to synthesize the thinking of meeting participants. Meeting summaries and documentation need not be as elaborate as the three-part product of the mini-conference, or even as lengthy as the meeting summary of the principal investigators' meeting. But a formal attempt needs to be made to draw conclusions based on some record or "evidence" of the meeting itself needs to be made so that the various meanings and imports of the meeting are available for later consideration.

Third, working seminars are most appropriate when there is something to be gained by the interchange of ideas or contrasting viewpoints. Many assessment questions--or questions about assessment approach--lend themselves to this kind of treatment, especially where there is significant disagreement and where appropriate experts can be assembled. Because science education lies at the intersection of many disciplines, assessment questions often raise such concerns.

Finally, the participation of NSF staff in the seminars themselves, as well as in planning them, is essential. If done well, working seminars can stimulate NSF staff to reflect about their investments in a way that is not easy in the normal course of their working lives at the Foundation. In so doing, NSF staff have another important opportunity to stay connected to the community of professionals that concern themselves with education in the sciences.

REFERENCES

Association of Science and Technology Centers. (In progress). *Survey of science museums.* Washington, DC: Author.

California State Department of Education. (1985). *Mathematics framework for California public schools, kindergarten through grade twelve.* Sacramento, CA: Author.

Conference Board of the Mathematical Sciences (CBMS). (1983). The mathematical sciences curriculum K-12: What is still fundamental and what is not. In *Educating Americans for the 21st century: Source materials.* Washington, DC: National Science Board.

Conference Board of the Mathematical Sciences (CBMS). (1984). *New goals for mathematical sciences education.* Washington, DC: Author.

Coxford, A. (1985). School algebra: What is still fundamental and what is not? In C. R. Hirsch & M. J. Zweng (Eds.), *The secondary school mathematics curriculum: 1985 yearbook of the NCTM.* Reston, VA: National Council of Teachers of Mathematics.

Crane, V. (1987). *An exploratory study of "3-2-1 Contact!" viewership.* Chestnut Hill, MA: Research Communications, Ltd.

Darling-Hammond, L., & Hudson, L. (1987). *Precollege science and mathematics teachers: Supply, demand and quality.* Washington, DC: Rand Corporation.

Elmore, R. (1980). *Complexity and control. What legislators and administrators can do about implementing public policy.* Washington, DC: National Institute of Education.

General Accounting Office. (1984). *New directions for federal programs to aid mathematics and science teaching.* Washington, DC: Author.

Greene, D., & David, J. L. (1984). A research design for generalizing from multiple case studies. *Evaluation and Program Planning, 7,* pp. 73-85.

House Appropriations Committee. (1987). *Committee report on HUD and independent agencies. Report 100-192.* Washington, DC: U.S. House of Representatives.

ISR Newsletter. How families use time. (Winter 1985-86, pp. 3-4.) Ann Arbor, MI: University of Michigan, Institute for Social Research.

Knapp, M. S. (1980). *Using the retrospective case history in exploratory organizational research: Research and planning unit development in a California community college district.* Menlo Park, CA: SRI International.

Knapp, M. S., Stearns, M. S., St. John, M., and Zucker, A. (1987a). *Opportunities for investment in K-12 science education: Options for the National Science Foundation--Summary report.* Menlo Park, CA: SRI International.

Knapp, M. S., St. John, M., Zucker, A. A., Needels, M., and Stearns, M. S. (1987b). *Opportunities for investment in K-12 science education: Options for the National Science Foundation. Volume 1: Problems and opportunities.* Menlo Park, CA: SRI International.

Knapp, M. S., Stearns, M. S., St. John, M., and Zucker, A. A. (1987c). *Opportunities for investment in K-12 science education: Options for the National Science Foundation. Volume 2: Groundwork for strategic investment.* Menlo Park, CA: SRI International.

Miles, M., & Huberman, M. (1984). *Qualitative data analysis: A sourcebook of new methods.* Beverly Hills, CA: Sage.

National Science Board. (1985). *Science indicators: The 1985 report.* Washington, DC: Author.

National Science Board. (1988). *Science and engineering indicators - 1987.* Washington, DC: National Science Foundation.

National Science Foundation. (1982). *Public understanding of science. Summary of grants and activities 1976-1981.* Washington, DC: Author.

National Science Foundation. (1986a). *Solicitation: Materials for elementary school mathematics instruction.* Washington, DC: Author.

National Science Foundation. (1986b). *Solicitation: Programs for preparing middle school science and mathematics teachers.* Washington, DC: Author.

National Science Foundation. (1987a). *Solicitation: Private sector partnerships to improve K-12 science and mathematics education.* Washington, DC: Author.

National Science Foundation. (1987b). *Program announcement: Office of studies and program assessment.* Washington, DC: Author.

National Science Foundation. (1987c). *RFP ESS 97-017: Data collection system for the Young Scholars Program.* Washington, DC: Author.

Oakes, J. (1987). *Opportunities, achievement, and choice: Issues in the participiation of women, minorities, and the disabled in science.* Prepared for the National Science Foundation.

Quick, S. K. (1977). *Secondary impacts of the curriculum reform movement: A longitudinal study of the incorporation of innovations of the curriculum reform mmovement into commercially developed curriculum programs.* Unpublished doctoral dissertation. Stanford, CA: Stanford University.

Romberg, T. A., & Stewart, D. M. (Eds.). (1984). *Proceedings of the Conference on School Mathematics: Options for the 1980s.* Washington, DC: U.S. Department of Education.

Senate Appropriations Committee. (1987, June 25). *Committee report on HUD and independent agencies. Report 100-189.* Washington, DC: U.S. Senate.

U.S. Department of Commerce. (1986). *Statistical abstract of the United States.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Education. (1987). *The condition of education.* Washington, DC: U.S. Government Printing Office.

Usiskin, Z. (1985). We need another revolution in secondary school mathematics. In C. R. Hirsch & M. J. Zweng (Eds.), *The secondary school mathematics curriculum: 1985 yearbook of the NCTM.* Reston, VA: National Council of Teachers of Mathematics.

Walberg, H. J. (1985). Synthesis of research on teaching. In M. C. Whittrock (Ed.), *The handbook of research on teaching. 3rd ed.* New York: Macmillan.

Weiss, I. R. (1988). *National survey of science and mathematics education.* Research Triangle Park, NC: Research Triangle Institute.

Welch, W. W. (1987). *Instructional content in mathematics and science. A synthesis of current policy issues.* Minneapolis: University of Minnesota.

Yin, R. K. (1984). *Case study research: Design and methods.* Beverly Hills, CA: Sage.

Appendix

ACKNOWLEDGMENTS

We owe a great deal to many individuals who helped us in completing Phase II of this study. Our special thanks go to the managers and staff of NSF's Science and Engineering Education Directorate. In addition to being the primary audience for this report, they were the source of many of its ideas, as well as a guide to the context in which any efforts to improve assessment must operate. In particular, we wish to acknowledge Merlyn Behr, Richard Berry, David Florio, Dorothy Gabel, Ray Hannapel, Alan Hoffer, Alan McClelland, Alice Moses, Charles Puglia, William Schmidt, Ethel Schultz, Susan Snyder, Arnold Strassenberg, Michael Templeton, George Tressel, and Robert Watson. We especially appreciate the openness and frankness these staff exhibited in their dealings with us throughout this project.

In carrying out the pilot test of short-term assessment activities, we relied on numerous individuals and groups in the informal science education community in addition to some of the NSF staff mentioned above. These people shared their thinking, data, or experiences in the various areas of informal science education addressed by pilot activities. We are especially grateful in this regard to Annette Berkovits, Margaret Cole, Robert Cook, Valerie Crane, Whitman Cross, Kay Davis, Elsa Feher, Frank Gardner, Keith Mielke, Roger Miles, Jon Miller, Philip and Phyllis Morrison, Vito Perrone, Wayne Ransome, Brett Waller, and Bernard Zubrovski, who participated in our two pilot working seminars.

In addition, we wish to thank Marilyn Eichinger, Paul Knappenburger, Patricia McNamara, Dennis Schatz, Max Suddeth, and other staff of museums participating in the Exhibit Research Collaborative, who shared with us their experiences in this project and their perspectives on the contributions NSF support had made to their work.

Various individuals reviewed or commented on our work in developing assessment approaches for NSF and they deserve mention as capable critics and shapers of some of our thinking. In particular, we note the contributions of Alphonse Buccino, Milbrey McLaughlin, Barbara Scott Nelson, and Mary Budd Rowe.

Finally, various members of the SRI project team played an indispensable role as data collectors, analysts, editors, and secretarial staff. To Carolyn Estey, Klaus Krause, Debra Shaver, Dorothy Stewart, Mark Stumbaugh, Joanne Taylor, Annette Tengan, Kathy Zacher, and Anita Smith (of Inverness Research Associates) we owe our gratitude for the unfailing professionalism and cheerfulness with which the many challenges of this project have been approached.

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025-3493
(415) 326-6200
TWX: 910-373-2046
Telex: 334 486